A nutrigenomics approach to understand the physiological properties of dairy products and probiotics

Author(s):
Sagaya, Mary Francina

Publication Date:
2011

Permanent Link:
https://doi.org/10.3929/ethz-a-006705916

Rights / License:
In Copyright - Non-Commercial Use Permitted
A nutrigenomics approach to understand the physiological properties of dairy products and probiotics

A dissertation submitted to
ETH Zurich

for the degree of
Doctor of Sciences

presented by
Sagaya Mary Francina
MSc Molecular Biology, University of Madras

born 16.10.1975
citizen of India

accepted on the recommendation of

Prof. Dr. Richard F. Hurrell, examiner
Prof. Dr. Walter Wahli, co-examiner
Dr. PD. Guy Vergères, co-examiner

2011
ACKNOWLEDGEMENT

I wish to express my sincere gratitude to all who have helped and encouraged me in different ways throughout this PhD, in particular:

Dr. Guy Vergères, my supervisor, for introducing me into this field of nutrigenomics, for his guidance, enthusiasm and continuous support throughout all these years and for always being available

Prof. Richard Hurrell, my supervisor and doctoral study advisor, for his continuous support, encouragement and inspiration, foremost always thinking about student’s best interest

I would like to thank Prof. Walter Wahli for accepting to be my thesis co-examiner.

I would like to thank people with whom I have collaborated during my PhD study. Dr. Michel Bellis, Dr. Werner Luginbühl, Dr. Frédéric Shütz, Dr. Ruan Elliott, Dr. Biljana Hacin, Dr. Bojana Matijasić.

ALP Colleagues: I would like to acknowledge my colleagues from ALP for making my PhD journey enjoyable and a learning experience. Dr. Barbara Walther, Dr. Robert Sieber, Alexandra Schmid, Karin Wegmüller, Doreen Gille from nutrition group. Brita Rehberger, Magali Chollet, Walter Strahm, Claude Hegel, Patrizia Piccinali, Katrin Bolanz, Flurina schwander, Dr. Hedwig Schlichterher-Cerny, Dr. Katharina Breme, Patrice Pittet, Yves, Ueli Butikofer, Brigitta, Marianne, Edith Rothlisberger.

A special thanks to Silvia Mallia and Katrin Bolanz for being great friends and office mates. Thank you Katrin for your help in german translation.

ETH lab colleagues: I want to thank all present and former colleagues at the Human Nutrition Laboratory at ETH Zürich for the good company and support: Dr. Isabelle aeberli, Dr. Maria Andersson, Jeannine Baumgartner, Raschida Bouhouch, Marica Bronic, Dr. Ralf Biebinger, Nico Brunner, Colin Cercamondi, Timo Christ, Dr. Eberhard Denk, Dr. Ines Egli, Maria Fischer, Valeria Galetti, Dominik Glinz, Dr. Stéphanie Good, Dr. Florentine Hilf, Dr. Matthias Hoppler, Dr. Karin Hotz, Martine Hurrell, Tanja Jaeggi, Adam Krzystek, Dr. Diego Moretti, Stefanie Murer, Nicolai Petry, Siwaporn Pinkaw, Sabine Renggli, Dr. Fabian Rohner, Käthe Santagata, Dr. Stefan Storcksdieck, Jasmin Tajeri Foman, Dr. Barbara Trösch, Dr. Thomas Walczyk, Dr. Rita Wegmüller and Christophe Zeder, Prof. Michael Zimmermann,

This work was possible through financial support from Agroscope Liebefeld-Posieux internal funding

Family: I thank my late parents for giving me all the support and encouragement throughout my education and also during the PhD. Thank you Akka, Prabu, Ezhil, Michael, Helen, Suja, Shristi, Kevin, Anita, Reena, Uncle & aunty and other family members for your support and encouragement.

Doing PhD, off course would have been much difficult without friends who always brought refreshing moments through their moral support, their enthusiasm to know about my research and its practical implication, and sharing their leisure hours having stimulating discussions, meals, outings etc. To mention a few, I would like to thank, Anita & Jordi, Blavna & Hemant, Chirdeep & Parminder, Christelle, Jahnvi & Jeet, Jasmine& Iqbal and Silvia & Tim.

Special thanks go to my God children Anahita and Reena for bringing a ray of hope in my life.

Last but most importantly, I would like to thank you!!! Mathew for all your help during my Phd period especially with this thesis, proofreading, correcting, for your motivation, understanding, cheerfulness and always diverting my attention and showing me the positive side to everything.
# TABLE OF CONTENTS

**ABBREVIATIONS** ............................................................................. 1  
**GLOSSARY** ..................................................................................... 3  
**SUMMARY** ..................................................................................... 5  
**ZUSAMMENFASSUNG** .................................................................. 9  
**INTRODUCTION** ............................................................................ 13  

## CHAPTER 1- LITERATURE REVIEW

1. **INTRODUCTION TO NUTRITIONAL GENOMICS** ................. 17  
   1.1 Diet as an environmental factor ............................................. 19  
   1.2 Ingestion, digestion and absorption of dietary components ...... 19  
   1.3 Mechanism of gene-nutrient interaction: nutrients as dietary signalling molecules .... 22  
   1.4 Nutritional Genomics ........................................................... 24  
      1.4.1 Genomics .......................................................................... 27  
      1.4.2 Transcriptomics ............................................................ 28  
      1.4.3 Proteomics ....................................................................... 29  
      1.4.4 Metabolomics ............................................................... 30  
      1.4.5 Systems biology ........................................................... 30  
      1.4.6 Bioinformatics .............................................................. 31  
   1.5 Benefits from nutrigenomics .................................................. 31  
   1.6 Summary .............................................................................. 33  
2. **MICROARRAY-BASED TRANSCRIPTOMICS** ...................... 34  
   2.1 Microarray technology ......................................................... 35  
      2.1.1 Microarray experiment .................................................. 38  
         2.1.1.1 Biological sample collection ................................. 39  
         2.1.1.2 RNA quality assessment .................................... 39  
         2.1.1.3 RNA amplification .............................................. 40  
         2.1.1.4 Target labeling .................................................... 41  
         2.1.1.5 Hybridization ...................................................... 41
2.1.1.6 Scanning ................................................................. 41
2.1.2 Microarray data analysis ........................................... 42
  2.1.2.1 Image analysis and quality assessment ................... 42
  2.1.2.2 Normalization .................................................... 43
    a) Global mean normalization .................................... 44
    b) Local normalization using Standardization and NOrmalization of MicroArray Data (SNOMAD) ........................................ 45
    c) Quantile normalization ......................................... 46
  2.1.2.3 Detecting differentially expressed genes .................. 47
  2.1.2.4 Data visualization for identifying patterns in microarray data ................................................................. 49
  2.1.2.5 Biological inference ............................................ 50
    a) Pathway enrichment analysis .................................. 51
    b) Transcription factor analysis .................................. 52
  2.1.2.6 Software for data analysis .................................... 56
  2.1.3 Microarray data depositories .................................. 57
  2.2 Application of transcriptomics in nutrition research ....... 57
    2.2.1 Transcriptomics studies using cell cultures ............. 57
    2.2.2 Transcriptomics studies using animal models .......... 58
    2.2.3 Transcriptomics studies in human ......................... 60
  2.3 Blood cell transcriptomics in humans .......................... 61
  2.4 Emerging technology in transcriptomics ....................... 64
  2.5 Summary .................................................................. 65

3. MILK AND DAIRY PRODUCTS ......................................... 66
  3.1 Milk components and their health effects ..................... 67
    3.1.1 Milk protein ....................................................... 67
    3.1.2 Milk lipid ............................................................ 68
    3.1.3 Lactose ............................................................... 69
    3.1.4 Minerals and vitamins ......................................... 70
    3.1.5 Microvesicles ..................................................... 70
  3.2 Yogurt components and their health effects ................. 71
  3.3 Metabolic and physiological effects of dairy products .... 72
    3.3.1 Metabolic effect of dairy components ..................... 73
    3.3.2 Anti-oxidative and anti-inflammatory effects of dairy components ................................................................. 75
    3.3.3 Cytomodulatory effects of dairy components .......... 80
3.3.4 Anti-hypertensive effects of dairy components..............................................84
3.4 Milk as carrier for probiotics............................................................................85
4. PROBIOTICS........................................................................................................87
  4.1 Mechanism of action of probiotics.................................................................89
    4.1.1 Influence of probiotics on the gut microflora........................................89
    4.1.2 Influence of probiotics on host intestinal epithelium, mucus production, and barrier function.................................................................91
    4.1.3 Influence of probiotics on immune responses.........................................91
  4.2 Transcriptomics in probiotic research............................................................94
  4.3 Lactobacillus gasseri K7 - a potential probiotic..............................................95

REFERENCES.........................................................................................................99

CHAPTER 2 – HUMAN STUDY..............................................................................123
CHAPTER 3 – TRANSCRIPTION FACTOR ANALYSIS......................................163
CHAPTER 4 – MOUSE STUDY.............................................................................183
CONCLUSION.....................................................................................................215
CURRICULUM VITAE.............................................................................................219
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>angiotensin converting enzyme</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARE</td>
<td>antioxidant response element</td>
</tr>
<tr>
<td>aRNA</td>
<td>antisense RNA</td>
</tr>
<tr>
<td>BAD</td>
<td>BCL2-associated agonist of cell death</td>
</tr>
<tr>
<td>BAG</td>
<td>BCL2-associated athanogene</td>
</tr>
<tr>
<td>BAK</td>
<td>BCL2-antagonist/killer</td>
</tr>
<tr>
<td>BAX</td>
<td>BCL2-associated X protein</td>
</tr>
<tr>
<td>BCAA</td>
<td>branched chain amino acid</td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma</td>
</tr>
<tr>
<td>BCL-10</td>
<td>B-cell CLL/lymphoma 10</td>
</tr>
<tr>
<td>BCL-(x,xL, xS, w)</td>
<td>BCL2-like</td>
</tr>
<tr>
<td>BID</td>
<td>BH3 interacting domain death agonist</td>
</tr>
<tr>
<td>BIK</td>
<td>BCL2-interacting killer</td>
</tr>
<tr>
<td>BIM</td>
<td>Bcl2-interacting mediator of cell death BCL2-like 11</td>
</tr>
<tr>
<td>BLK</td>
<td>B lymphoid tyrosine kinase</td>
</tr>
<tr>
<td>CPP</td>
<td>caseinophosphopeptide</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>ChREBP</td>
<td>carbohydrate response element binding protein</td>
</tr>
<tr>
<td>ChRE</td>
<td>carbohydrate response element</td>
</tr>
<tr>
<td>CLOVER</td>
<td>Cis-eLement OVERrepresentation</td>
</tr>
<tr>
<td>DD</td>
<td>death domain</td>
</tr>
<tr>
<td>EAA</td>
<td>essential amino acid</td>
</tr>
<tr>
<td>EGF</td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td>FADD</td>
<td>Fas-associated death domain</td>
</tr>
<tr>
<td>FC</td>
<td>fold change</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery change</td>
</tr>
<tr>
<td>FGF</td>
<td>fibroblast growth factor</td>
</tr>
<tr>
<td>Glut-2</td>
<td>glucose transporter 2</td>
</tr>
<tr>
<td>Glut-5</td>
<td>glucose transporter 5 (fructose transporter)</td>
</tr>
<tr>
<td>GSEA</td>
<td>gene set enrichment analysis</td>
</tr>
<tr>
<td>HGP</td>
<td>human genome project</td>
</tr>
<tr>
<td>IAP</td>
<td>inhibitor of apoptosis</td>
</tr>
<tr>
<td>IGF</td>
<td>insulin growth factor</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto encyclopedia for genes and genomes</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>NF-kB</td>
<td>nuclear factor 'kappa-light-chain-enhancer' of activated B-cells</td>
</tr>
<tr>
<td>NRF2</td>
<td>NF-E2-related factor 2</td>
</tr>
</tbody>
</table>
MGED  microarray gene expression data
MIAME  minimum information about a microarray experiment
MFGM  milk fat globule membrane
MM  mismatch
mTOR  mammalian target of rapamycin
PCA  principle component analysis
PEPT1  peptide transporter 1
PPARs  peroxisomes proliferator activated receptors
PM  perfect match
PUFA  polyunsaturated fatty acid
RAR  retinoic acid receptor
RIN  RNA integrity number
ROS  reactive oxygen species
SFA  saturated fatty acid
SGLT1  sodium dependent glucose transporter 1
SNP  single nucleotide polymorphism
TGF  transforming growth factor
TIFF  tagged image file format
TLR  toll like receptor
TNF  tumor necrosis factor
UHT  ultra high temperature
VDR  vitamin D receptor
GLOSSARY

Allele - any one of a number of alternative forms of the same gene occupying a given position on a chromosome.

Bioactive components - bioactive components are compounds occurring in food that bring about a physiological response/effect in an organism.

cDNA - Complementary DNA; single-stranded DNA that is complementary to mRNA in the presence of reverse transcriptase.

Database - a database is a system intended to organize, store, and retrieve large amounts of data easily. It consists of an organized collection of data for one or more uses, typically in digital form.

Data mining - a branch of computer science, is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence.

Down-regulation - a negative regulatory effect on a physiological process. At a molecular level, the regulatory sites include membrane receptors, gene, messenger RNA, and proteins.

Feature - spot containing single stranded DNA/cDNA/oligonucleotides on the array.

Feature intensity – intensity of the spot after hybridization.

Fabrication - method of placing the probes on the physical surface of the array/chip.

Gene - the fundamental physical and functional unit of heredity, which carries information from one generation to the next; a segment of DNA, composed of a transcribed region and regulatory sequences that make transcription possible.

Genomics - the study of genome which is nothing but all the genetic material in the chromosome of a particular organism. Its size is generally given as its total number of base pairs.

Genotype - the genetic constitution of an organism, as distinguished from its physical appearance (its phenotype).

Gene ontology - the Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases using controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data.

Homeostasis - homeostasis is a property of system or any self-regulating process by which biological systems tend to maintain stability while adjusting to conditions that are optimal for survival.

Human genome project - Human genome project is an international collaboration that successfully provided researchers with basic information about the sequence of the three billion chemical base pairs that make up human genomic DNA.

Metabolomics - the study of metabolome (set of metabolic entities and small pathway motifs in a cell, tissue, organ, organism, and species) produced by an organism at any given time.

Mismatch probes - one nucleotide is a mismatch from the gene sequence.

mRNA - RNA, synthesized from a DNA template during transcription that mediates the transfer of genetic information from the cell nucleus to ribosomes in the cytoplasm, where it serves as a template for protein synthesis. Also called messenger RNA.
Nutrigenetics - aims to understand how the genetic makeup of an individual coordinates their response to the diet, and thus considers underlying genetic polymorphisms.

Nutrigenomics - aims to determine the influence of common dietary ingredients on the genome, and attempts to relate the resulting different phenotypes to differences in the cellular response of the biological system.

Nutritional preemption - process of disease risk reduction via nutritional intervention approaches through a greater understanding of nutrigenomics, proteomics and metabolomics.

Perfect match probes - perfectly matches the gene sequence.

Phenotype - the visible properties of an organism that are produced by the interaction of the genotype and the environment.

Photolithography - technique used by affymetrix for the fabrication of probes.

Polymorphism - existence of a gene in a several allelic forms.

Prebiotics – nondigestible food components that support overall health by promoting the activity of probiotic bacteria in the large intestine.

Probiotics - probiotics, as defined by World Health Organization, are live micro-organisms which, when administered in adequate amounts, confers a beneficial effect on the health status of the host.

Probes - probe is a single stranded sequence of DNA/cDNA on a microarray slide.

Promoter - in genetics, a promoter is a region of DNA that facilitates the transcription of a particular gene. Promoters are located near the genes they regulate, on the same strand and typically upstream (towards the 5’ region of the sense strand).

Proteomics - the study of proteome (a set of proteins produced by an organism at any given time) of an organism.

Transcriptomics - the study of the transcriptome, (a subset of genes transcribed in a given organism), the complete set of RNA transcripts produced by the genome at any one time.

Transcription factor - transcription factors are proteins that regulate or control the activity or expression of genes.

Up-regulation - a positive regulatory effect on a physiological process. At a molecular level, the regulatory sites include membrane receptors, genes, messenger RNA and proteins.
SUMMARY

**Background:** The Human Genome project has not only paved the road to gaining information on the human genome but has also provided a complete set of high-throughput techniques for elucidating the genome of further species and for a global characterization of further molecules along the molecular chain of cellular information namely mRNA, proteins and metabolites. Together, genome sequencing projects and high-throughput techniques has the potential to induce a paradigm shift in nutrition research. Along this line, nutrigenomics is a new branch of nutrition research that allows one to study the interaction of food with living organisms at the level of gene, protein, and metabolite in a holistic manner.

Dairy products have become part of western diet. There is a growing interest in understanding the potential health benefits of milk and milk-based products, such as yogurt and probiotics. However, there are studies that challenge such notions. Nutrigenomics is an analytical tool that could help in resolving it by elucidating the molecular events that are induced by milk and dairy-based products.

**Aim of this thesis:** The objective of this thesis was to exploit the potential of nutrigenomic technologies to investigate how dairy products and probiotics affect physiological functions *in vivo* by altering the gene expression in whole blood cells.

**Experiments:** This objective was addressed by conducting two independent studies which are described and discussed in this thesis.

In the human study, the genome-wide postprandial effects of the ingestion of milk and yogurt by healthy human subjects were investigated using blood cell transcriptomics. A randomized, controlled, single blinded, crossover study on 6 healthy male individuals was carried out. After an overnight fast, 540 grams of milk or yogurt was ingested by the subjects. Blood samples were collected before (0h) and 2h, 4h, 6h after ingestion and the blood cell transcriptome was analyzed using a linear kinetic analysis.

In the mouse study, the anti-microbial effects of a potential probiotic strain of lactic acid bacteria (*Lactobacillus gasseri* K7 (Rif³)) were investigated in mice challenged with a gastrointestinal enterohemorrhagic pathogen (*Escherichia coli* O157:H7). The blood cell
transcriptomes of five groups of mice were measured: (1) control C57BL/6J mice; (2) mice inoculated with enterohemorrhagic *E. coli* O157:H7; (3) mice inoculated with *L. gasseri* K7(Rif²); (4) mice pre-inoculated with *L. gasseri* K7(Rif²) and subsequently inoculated with *E. coli* O157:H7 (5) mice co-inoculated with *E. coli* O157:H7 and *L. gasseri* K7(Rif²).

**Result and discussion:** Dairy ingestion by healthy human subjects revealed metabolic and immunomodulatory responses in their blood cells. The expression of genes in these two groups showed a biphasic kinetic trend. The genes in the metabolic group first showed a decrease in expression when compared to control between 0h and 2h which was transient before being more profoundly up-regulated between 2h and 6h. This group is mostly composed of genes involved in protein synthesis and mitochondrial function. Increased gene expression in this group may reflect the metabolic response of the organism to the ingestion of the macronutrients present in the dairy products. The increased protein synthesis response may also reflect the increased metabolic activity of the peripheral blood cells, especially leukocytes. The expression of the second group of genes corresponding to immunomodulatory group showed first a transient increase between 0h and 2h before being more markedly down-regulated between 2h and 6h. This group is mostly composed of genes involved in immunomodulatory processes such as inflammation and apoptosis. In particular, down-regulation of inflammatory pathways, such as the Toll-like receptor/NF-κB pathway, between 2h and 6h is in line with studies in humans and animal models documenting anti-inflammatory properties of dairy products. Furthermore, an analysis of the transcription factor network that triggers the observed gene expression profile revealed, among others, an enrichment of the transcription factor NRF2 a finding that gives further clues on the mechanisms of action of dairy products. Finally, a comparison of the gene expression profiles resulting from the ingestion of milk and yogurt revealed eight pathways in the yogurt group that are significantly different from the milk group. Among these, the ACE2 pathway was modulated, an interesting finding in light of the recent interest devoted to the production of anti-hypertensive peptides derived from the fermentation of casein in dairy products.

No clinical signs of infection in the *E. coli* O157:H7 group of mice were observed. However, the blood cell transcriptome could clearly distinguish between the five treatment groups. In the *E. coli* O157:H7 group 1’541 genes were differentially expressed and a functional analysis revealed changes in the expression of genes that regulate pathways characteristic of bacterial infection (cell adhesion, cytoskeleton rearrangement, inflammation, and glycolysis).
Furthermore, inoculation with *L. gasseri* K7(Rif') suppressed the gene expression profile induced by *E. coli* O157:H7, the effect being more pronounced in the pre-incubation than in the co-incubation protocol.

**Conclusion:** Several human intervention studies have now demonstrated that blood cells can reprogram their cellular status by inducing/modulating their gene expression profile in response to changes in their nutritional environment. The work in this thesis demonstrates that blood cells do respond to dietary components present in dairy products, in particular to milk and yogurt. In addition, it also demonstrates that blood cell transcriptomics can be a sensitive tool to monitor interaction between the host, probiotics and pathogens.

In the nutritional intervention study, the crossover design and kinetic analysis of the postprandial blood cell transcriptome of healthy human subjects having ingested dairy products allowed the identification of statistically significant and biologically relevant processes. Three comments highlight the importance of multiple time points measurement of gene expression patterns: (a) Gene expression is a dynamic state, (b) gene expression kinetics are not necessarily monotonous, (c) gene expression induced by dietary components are subtle and rarely involve high fold changes. Thus, analyzing the kinetics of gene expression (up- and down-regulating) can be more robust and informative than simple observation including two time points (e.g. before and after). From the biological point of view, the decreased expression of genes related to inflammation at 6h compared to 0h suggests a potential role for dairy products in the management of chronic inflammatory diseases. In addition, the enriched presence of binding sites for the transcription factor NRF2 in genes coding for protein synthesis and mitochondrial function suggests that a postprandial regulation of the anti-oxidant system is taking place after the ingestion of dairy products. Thus, data obtained in this study underline the possibility that an anti-inflammatory response may be induced after the ingestion of dairy products by activating the NRF2-mediated antioxidant machinery as well as by suppressing pro-inflammatory pathways mediated by NF-κB signalling.

Under suboptimal experimental conditions in which the inoculated dose of *E. coli* is not sufficient to induce clinical signs of pathogenicity, the blood cell transcriptome of mice allows to differentiate between the various treatments, suggesting that blood cell transcriptomics is a sensitive analytical strategy to evaluate the potential probiotic function of strains of lactic acid bacteria in vivo. At term, such studies will eventually allow the
identification of biomarkers for the selection of bacteria transforming milk into products with enhanced nutritional properties.
ZUSAMMENFASSUNG


In der Studie mit Mäusen wurden die anti-mikrobiellen Effekte eines potentiell probiotischen Stamms Milchsäurebakterien Lactobacillus gasseri K7 (Rif6) in Mäusen untersucht, welche mit einem gastrointestinalen, enterohämorrhagischen Pathogen (Escherichia coli O157:H7) infiziert wurden. Die Blutzelltranskriptome von fünf Gruppen von Mäusen wurden gemessen: (1) Kontrollmäuse C57BL/6J; (2) Mäuse angeimpft mit enterohämorrhagischen E. coli O157:H7; (3) Mäuse angeimpft mit L. gasseri K7(Rif6); (4) Mäuse, welche vorher mit L. gasseri K7(Rif6) angeimpft wurden und anschließend mit E. coli O157:H7; (5) Mäuse gleichzeitig angeimpft mit E. coli O157:H7 und L. gasseri K7(Rif6).

Interesse an der Produktion von Blutdruck senkenden Peptiden, die von der Fermentation von Kasein in Milchprodukten stammen.

Es wurden keine klinischen Zeichen einer Infektion in der E. coli O157:H7 Gruppe von Mäusen beobachtet. Mittel Blutzellentranskriptomik konnten die fünf Mäusegruppen jedoch klar unterschieden werden. In der E. coli O157:H7 Gruppe wurden 1’541 Gene unterschiedlich exprimiert welche Stoffwechselwege regulieren, die karakteristisch für bakterielle Infektionen sind (Zelladhäsion, Cytoskeletnanordnung, Inflammation, Glykolyse). Desweiteren unterdrückte eine Beimpfung mit L. gasseri K7(Rif') die Genexpressionsprofile, die durch E. coli O157:H7 induziert wurden, wobei der Effekt bei der vorhergehenden Inkubation stärker war als bei der gleichzeitigen Inkubation.


Das überkreuzte Design und die kinetische Analyse des postprandialen Blutzellentranskriptoms in der Ernährungsinterventionsstudie mit gesunden Menschen, ermöglichte die Identifikation von statistisch signifikanten und biologisch relevanten Prozessen. Drei Erläuterungen betonen die Bedeutung von multiplen Zeitpunktmessungen von Genexpressionsmustern: (a) Die Genexpression ist in einem dynamischen Zustand, (b) Genexpressionskinetiken sind nicht unbedingt monoton, (c) eine durch Nahrungskomponenten induzierte Genexpression sind sehr empfindlich eingestellt und zeigt selten grösse Änderungen. Deshalb sollte eine Analyse der Kinetik der Genexpression (Auf- und Abregulation) robust und informative sein als eine einfache Beobachtung, welche nur zwei Zeitpunkte enthält (z.B. voher und nachher). Vom biologischen Standpunkt aus suggeriert die heruntergeregelte Expression von Genen, die mit Inflammation in Verbindung stehen nach 6h im Vergleich zu 0h eine potentielle Rolle von Milchprodukten in der Modulation von chronischen Inflammationskrankheiten. Zusätzlich suggeriert die erhöhte Anwesenheit von Bindungsstellen für den Transkriptionsfaktor NRF2 in Genen, welche für die Proteinsynthese und Funktion der Mitochondrien kodieren, dass eine postprandiale
Regulation des antioxidativen Systems stattfindet nach dem Verzehr von Milchprodukten, Folglich unterstreichen die Daten die in dieser Studie erhalten wurden die Möglichkeit, dass nach dem Verzehr von Milchprodukten eine anti-inflammatorische Antwort induziert werden könnte durch die Aktivierung eines von NRF2-vermittelten antioxidativen Mechanismus wie auch durch eine Unterdrückung von durch NF-κB Signale vermittelten pro-inflammatorischen Stoffwechselwegen.

INTRODUCTION

Diet is an important environmental factor influencing our health. Modern nutrition research focuses on promotion of health, prevention of disease, and on performance improvement and risk assessment (Kussman and Blum, 2007; Trujillo et al., 2006). Understanding how diet affects the human health and to prevent the diet-related diseases, requires the understanding of how nutrients act at the molecular level. This would involve studying the interaction of common food/dietary components (nutrients and non-nutrients) at the level of genes, proteins and metabolites. A new branch of nutrition research has emerged that applies new high-throughput profiling techniques in order to better understand the interplay between genome and its nutritional environment at the level of transcripts, proteins and metabolites (Wittwer et al., 2011). This new branch in nutritional science is referred to as nutritional genomics.

Transcriptomics or transcript profiling utilizes high-throughput techniques to understand the interaction between the food/dietary components and genes (Corthesy-Theulaz et al., 2005). Microarray technique is one such high throughput technique now increasingly used for gene expression profiling. Its benefit lies on the fact that it gives a holistic view on the molecular events taking place as a result of interactions between food/dietary components and genes. Microarray-based nutritional genomic studies require access to transcriptionally active cells/tissue from which the influence of diet on genotype can be sensed effectively.

Transcript profiling using blood cells is developing into a mainstream tool for the assessment of the interactions between food/dietary components and genes. Indeed, various studies clearly show that blood cells respond to dietary components by inducing specific gene expression profiles (Bouwens et al., 2009; van Erk, Blom, van Ommen & Hendriks, 2006). Blood cells can thus be a useful less invasive method to monitor the effects of dietary interventions in human subjects. So far the blood cell transcriptomics studies have been more observational in nature (Wittwer et al., 2011). In this thesis we have applied the blood cell transcriptomics approach as proof of concept to understand the effect of dairy products and the potential probiotic strain *Lactobacillus gasseri* K7 on gene expression.

As dairy products, especially milk and yogurt, form an important part of our diet and are associated with health benefits (Haug, Hostmark & Harstad, 2007; Mills, Ross, Hill, Fitzgerald & Stanton, 2011), it becomes important to understand the molecular mechanism by
which they induce such effects. In that context, blood cell transcriptomics provide a useful analytical tool to monitor the molecular events.

The present thesis consists of a literature review followed by three manuscripts. Nutritional genomics is a multi-disciplinary field and requires the involvement of several disciplines, such as molecular biology, statistics, nutrition, physiology and bioinformatics. The literature survey briefly reviews these aspects. The literature review is divided into 4 sections, namely introduction to nutritional genomics; microarray-based transcriptomics and its application; a brief overview on milk and yogurt components and their effects on health, and finally a brief overview on probiotics and their properties. The first manuscript describes a blood cell transcriptomics approach to study the effect of milk and yogurt ingestion in healthy human subjects. The second manuscript further investigates the potential regulatory factors (transcription factors) responsible for the observed gene expression patterns induced by dairy ingestion. In the third manuscript, the antimicrobial properties of a potential probiotic strain, *Lactobacillus gasseri* K7 have been investigated using the blood cell transcriptome.
CHAPTER 1 - LITERATURE REVIEW
1. INTRODUCTION TO NUTRITIONAL GENOMICS

Understanding the relationship between food/diet and the human body is a challenging task and at the same time it is very important for our wellbeing.

Our understanding about the interaction between food and the human body has evolved over time (as illustrated in Figure 1) with the development and contribution of different disciplines in sciences, mainly chemistry, biochemistry, cell biology, physiology, epidemiology, genetics and molecular biology (Fairweather-Tait, 2003; Mehrotra, 2006; Ordovas & Corella, 2004).

![Timeline of Nutritional Genomics](image)

*Figure 1* Timeline in the contribution of diverse sciences to the increased knowledge and understanding of food and its role in the human body.

To understand the interaction between food and human body one starts with a phenotype which is an observable characteristic or trait of an organism. The phenotype can result from the influence of genotype (genetic constituent of an organism) or environmental factors, or both (Figure 2).
Figure 2 Illustration of phenotype-genotype-environment (diet) interactions.

Common environmental factors that affect the phenotype include elements of lifestyle (e.g., diet, smoking, alcohol consumption), socio-economic status and environment such as air pollution, microorganisms (Ordovas & Corella, 2004). The diet or food we consume is one of the main environmental factors regularly challenging the human biological system.

Traditional nutrition sciences have focused mainly on the interaction between environmental factors and phenotype, in particular on nutrient deficiencies and their effects on health and disease (Kussmann, Raymond & Affolter, 2006). Relatively less attention has been given towards the “genomics” aspect. Over the past few decades, there has been an increase in diet-related diseases, such as obesity, type 2 diabetes, cardiovascular disease (Afman & Muller, 2006; Fairweather-Tait, 2003; Milner, 2007). Genetic predisposition can be an important contributor to the development of diet-related diseases (Afman & Muller, 2006; Muller & Kersten, 2003). In addition, different components of food/diet can influence the metabolic and physiological processes/activities in an organism and play an important role in the control of homeostasis. As a result, there is a growing interest in determining the role of food/diet and its interaction with the human biological system at different molecular levels. In other words, studying the interaction between the genotype and environmental factors (food/diet) and its influence on the phenotype with the goal of preventing the development of
diseases using nutrition as a preventive measure. In this regard, nutritional genomics has emerged as a new approach in nutrition science, an influence that has been facilitated by the completion of the human genome project (Human Genome Sequencing Consortium, 2004) and the development of high-throughput technologies referred to as “OMICS” tools.

1.1 Diet as an environmental factor

Food/diet is a complex mixture of hundreds of different components. Humans are exposed to this complex mixture throughout their lifetime, and that makes diet one of the important environmental factor challenging the human biological system regularly (Ordavos & Corella, 2004). The human diet is mainly composed of macronutrients such as, carbohydrates, proteins, fat, and micronutrients, such as vitamins and minerals. The main sources for food products are plants and animals. Animal products, vegetables, and fruits in addition to the essential nutrients such as, essential amino acids, essential fatty acids, calcium, zinc, selenium, folate, vitamins C, and E also contain a variety of non-essential nutrients and bioactive components (Trujillo, Davis & Milner, 2006; Milner, 2007). Bioactive components are of special interest as they are known to affect biological processes and hence have an impact on body function or condition and ultimately on health (Korhonen, 2005). Indeed, many non-essential bioactive components promote healthy effects, and thus are being used as medicines (McKay & Blumberg, 2006; Korhonen, 2005). Overall, the dietary components found in food matrices can act in synergistic or antagonistic way in order to induce a physiological response in the human body (Lopez-Exposito, Pellegrini, Amigo & Recio, 2008). Thus, understanding the complex interplay between hundreds of absorbed food components and their influence in regulating various cellular processes could be important to define a diet as health-promoting or disease-promoting.

1.2 Ingestion, digestion and absorption of dietary components

As described earlier, food is a complex mixture of macronutrients and micronutrients. Food is accessible to living cells only when various complex food components are broken down into much smaller units. The process of ingestion, digestion, and absorption (involving multiple steps) facilitates this process. The gastrointestinal tract is the site where assimilation of nutrients and elimination of waste products takes place.

The mechanical and enzymatic digestion of food starts in the mouth, where saliva acts as a solvent for solid foods. Enzymes such as α-amylase from the salivary gland contribute to the
initial hydrolysis of the ingested food. The moistened and partly hydrolyzed food from the mouth is then passed on to the stomach where food is mixed with gastric juice resulting in chyme formation. The chyme from the stomach is then released into the upper part of the small intestine, namely the duodenum. The epithelial cells lining the small intestine contribute to the production of digestive enzymes. In response to chyme passing into the duodenum, bile and pancreatic juice are secreted into the duodenal lumen. The enzymatic secretion contains a variety of enzymes for digestion of carbohydrates (α-amylase), proteins (trypsin, chymotrypsin, carboxypeptidase and elastase), and lipids (lipase, phospholipase, esterase). Most of the digestion and absorption of nutrients takes place in the small intestine (Kutchai, 1998).

The majority of the ingested carbohydrate is quickly hydrolyzed to di- and mono-saccharides. Di- and oligo-saccharides are further digested by enzymes produced by the epithelial cells (enterocytes) of the duodenum and jejunum followed by absorption of monosaccharides.

Dietary proteins are digested partly by pepsin in the stomach and partly by proteases secreted by the pancreas (trypsin, chymotrypsin, carboxypeptidase and elastase). The digestive enzymes reduce the proteins to peptides which are further digested into amino acids and di-, tri- and tetra-peptides by membrane integrated peptidases produced by the epithelial cells of the duodenum and jejunum. The amino acids and small peptides are then absorbed and released into the blood.

Digestion of dietary fats is a complex process involving emulsification and micelle formation and lipolysis. The primary dietary lipids are triglycerides, which are emulsified in the small intestine with the help of bile acids. This produces emulsion droplets allowing the access of water-soluble lipolytic enzymes produced by the pancreas. In order to reach the epithelial surface, the digestion products (free fatty acids and mono-glycerides) need to form micelles. Bile acids due to their amphipathic structure (hydrophilic and hydrophobic) are capable of forming micelles and to carry lipids to the surface of the epithelium. Here, the micelles are disrupted and the lipids diffuse into the epithelial cells where they are packed as chylomicrons. From the epithelial cells chylomicrons enter the lymphatic duct for transport and eventually enter the blood stream (Goodman, 2010).

Figure 3 illustrates the digestion and absorption of macronutrients and the transport to blood.
The gastrointestinal tract is also home for about 100 trillion bacteria. The microbiota present in the gut plays an important role in metabolism of digested and non-digested food components (O’Hara & Shanahan, 2006; Xu et al., 2003; Musso, Gambino & Cassader, 2010). As a result, the intestine has become a target of a number of studies which aims at characterizing both the normal functional state of the tissue and its response to nutrients, and response to beneficial microorganisms (such as probiotics) and pathogens (Sonnenburg, Chen & Gordon, 2006; Roberts, Mutch & German, 2001).
1.3 Mechanism of gene-nutrient interaction: nutrients as dietary signalling molecules

Cells are constantly assessing and responding to environmental changes, thereby maintaining a regulated system that deals with the presence of nutrients or food/dietary components, pathogens or toxic components resulting from the absorption process. The absorbed food/dietary components behave as informational molecules (i.e. act as dietary signals). These dietary signals modulate the activity of intracellular proteins called transcription factors either directly or indirectly to regulate gene expression (Muller & Kersten, 2003). Figure 4 illustrates the direct and indirect mode of regulation of gene expression by nutrients.

**Figure 4** Nutrients as signalling molecule regulating the intracellular adaptation in a cell. Dark blue solid arrows show the direct regulation of gene expression by nutrients. Green dotted arrows show the indirect regulation of gene expression by a nutrient. Thin black arrows show the genetic information flow from DNA to RNA, protein and metabolites to finally regulation of biological function.

In case of direct gene regulation the nutrients bind to transcription factors. A family of transcription factor which directly senses the presence of nutrients is the nuclear receptor superfamily. This family consists of 48 members and is the most important group of nutrient sensors. Numerous receptors in this family bind nutrients and their metabolites. For example,
the retinoic acid receptor (RAR) binds retinoic acid, the vitamin D receptor (VDR) binds vitamin D, and the peroxisome proliferator activated receptors (PPARs) bind fatty acids. The interaction between nutrients and transcription factors leads to conformational changes in transcription factors. This change in structure facilitates the binding of transcription factors to a specific region (promoter regions) in DNA segment and thereby regulates gene expression (Afman & Muller, 2006; Corthesy-Theulaz et al., 2005; Desvergne, Michalik & Wahli, 2006; Muller & Kersten, 2003). An example of nutrient-induced gene expression is the activity of fatty acids as signalling molecules and their influence on gene expression in hepatic cells. Fatty acids are energy-rich molecules. During fasting, when energy is required by the cells, the energy-rich fatty acids stored in adipose tissues are released. Fatty acids are then transported to the liver. In the liver cells, the nuclear receptor PPAR member (PPARβ/δ) senses the presence of fatty acid and induces the expression of genes by binding to PPAR response elements in the DNA. The expressed genes are then translated to proteins that carry out the processes such as fatty-acid oxidation and ketogenesis to release energy (Muller & Kersten, 2003; Sanderson et al., 2009).

In case of indirect regulation of gene expression, nutrients influence intracellular signalling pathways, and/or the secretion of hormones or metabolites that in turn modulate the activity of transcription factors. An example of indirect regulation of gene expression comes from the regulation of gene expression by glucose. Intake of a carbohydrate-rich diet induces an increase in glucose concentration in plasma, in addition to the changes in concentration of insulin and glucagon. Glucose cannot directly regulate gene expression but its metabolites (glucose-6-phosphate and xylulose-5-phosphate) act as signalling molecules that in turn activate the transcription factor carbohydrate response element binding protein (ChREBP). ChREBP binds to carbohydrate response element (ChRE) in a specific region in DNA and stimulates the transcription of glucose-responsive genes (Corthesy-Theulaz et al., 2005).

The literature has widely documented that dietary components, such as carbohydrates, proteins, fatty acids, zinc and iron participate in the regulation of gene expression (Desvergne, Michalik & Wahli; 2006; Elliott, 2008). In short, nutrients are able to influence a wide array of specific genes and cellular functions by activating transcription factors (Müller & Kersten 2003; Corthésy-Theulaz et al., 2005; Afman & Müller 2006).

The following section describes the high-throughput technologies used to measure the interactions of food or nutrients at different molecular levels.
1.4 Nutritional Genomics

Two main factors contributed in the development of the so called “OMICS” era. First it was the “Human genome project” (HGP) and secondly the development and use of high throughput molecular techniques developed for the purpose of speeding up HGP. HGP is a worldwide effort to sequence the entire human genome with the aim of finding all the genes and their function in humans. Sequencing of the human genome was completed in 2003. This resulted in an explosion of information and the realization that a holistic perspective of the human biological system can be obtained via large scale studies addressing molecular interactions (Kaput & Rodriguez, 2004).

The term nutritional genomics was originally defined as “the general approach to gene discovery that is currently most applicable to compounds of nutritional importance that are synthesized or accumulated by plants and other organisms” (DellaPenna, 1999). Currently, nutritional genomics encompasses the study of the genome-wide influence of nutrition with the goal to better understand how nutrition influences metabolic pathways and homeostatic control, how this regulation is altered in the early phase of diet-related diseases, and to what extent individual sensitizing genotypes contribute to such disease (Muller & Kersten, 2003; Ordovas et al., 2004). Alternately, nutritional genomics aims at using diet to prevent or treat disease by studying the functional interaction of food and its components with the genome at the molecular, cellular, and systemic level.
The concept of nutritional genomics (Kaput & Rodriguez, 2004; Trujillo, Davis & Milner, 2006) builds on the premises that

1. Food components can behave as signalling molecules and act on the human genome, either directly or indirectly, to alter the expression of genes and gene products.
2. Diet can potentially compensate for or accentuate the effects of genetic polymorphisms.
3. The consequences of a diet are dependent on the balance between health and disease states and on the individual’s genetic background.
4. Diet and dietary components can alter the risk of disease development by modulating multiple processes involved in the onset, incidence, progression, and/or severity.

Nutritional genomics can be further classified into two disciplines (Mutch, Wahli & Williamson, 2005), namely
1. **Nutrigenomics** aims to understand the influence of food/dietary components on the genome of an organism, and attempts to relate the resulting different phenotypes to differences in the cellular response of the biological system.

2. **Nutrigenetics** aims to understand how the genetic makeup of an individual influences their response to the diet, and thus considers underlying genetic polymorphisms.

Both nutrigenomics and nutrigenetics embodies the science of identifying and characterizing genes associated with specific responses to food/dietary components and relating specific variation in this response to the maintenance of health and the development of disease states. Therefore, both disciplines aim to unravel diet/genome interactions; however, their approaches and immediate goals are distinct (Mutch, Wahli & Williamson, 2005).

To better understand the distinction between nutrigenomics and nutrigenetics let us consider the following analogy. A shirt or pant is made of fabric. A “generic” readymade shirt or pant (of different cloth materials) that fits to a general population based on shoulder size or hip size can be bought in shops such as a supermarket. On the other side, an individual could buy a piece of fabric or a readymade shirt or pant, go to a tailor (who takes different measurements in addition to shoulder size or hip size), and get stitched or adjusted so that the shirt or pant “perfectly” fits him/her. Nutrigenomics can be compared to the first scenario of readymade clothes, where the immediate goal is to unravel the optimal diet for a general population from within a series of nutritional alternatives. While, nutrigenetics can be compared to the latter scenario of tailor made cloth where the immediate goal is to identify the optimal diet for a given individual, i.e. personalized nutrition. In this thesis, we have performed nutrigenomic studies. In the later parts of this section we therefore mainly focus our discussion on nutrigenomics. For more literature on nutrigenetics the reader is referred to (Simopoulos, 2010; Rimbach & Minihane, 2009).

As described earlier, food/dietary components can influence genes and cellular processes by modulating transcription factor activity. This influence can be observed at different molecular levels, such as gene expression, protein expression, or metabolites production. Different omics technologies (transcriptomics, proteomics, metabolomics) are available, that help unravelling the influence of food/dietary components at different molecular levels. A nutritional genomics study can focus either on only one of the molecular levels or on all of
them together. In the latter case, it is referred to as systems biology. We briefly discuss each
of the omics technologies as well as systems biology below.

The human genome project has provided new tools (the so-called post-genomic high-
throughput technologies) and information on the human genetic code (genome). Figure 5
illustrates the different post-genomics technologies that can be applied in nutrition research to
obtain a more holistic perspective on how diet interacts with human biological system.

1.4.1 Genomics

Genomics is the study of mapping, sequencing, and analysing DNA, the genetic material
present in an organism. The genome of an organism contains all nucleotide sequences
including structural genes, regulatory sequences, and non-coding DNA sequences.

The human genome is estimated to encode 20'000 to 25'000 genes (Human Genome
Sequencing Consortium 2004). The vast amount of sequences and information from HGP has
also led to a better understanding of the human diversity (Venter et al., 2001). Organisms
such as humans are quite similar at the genetic level. However, differences exist at a
frequency of about 1 in every 1000 nucleotide bases, which translates into approximately 3
million base differences between each individual out of 3 billion base pairs. Such differences
are referred to as single nucleotide polymorphisms (SNPs). A significant effort is going on in
the research community (international hapmap project and human variome project) to map
the individual SNPs in humans and other organisms (Iafrate et al., 2004; Frazer et al., 2007;
Altshuler et al., 2010; Horaitis & Cotton, 2004; Cotton, Vihinen & den Dunnen, 2011). One
of the strategies that is being used to draw associations between diseases, genes, and nutrients
is the examination of haplotypes or haplotype blocks. A haplotype is the set of SNP alleles
along a region of a chromosome. Haplotype analysis can be used to identify groups of SNPs
linked together, and therefore may be useful in understanding the distribution of risk alleles
in human populations as well as for tailoring prevention strategies to those at increased risk
(Sabeti et al., 2007). An example of haplotype diversity is the lactase gene in humans.

Some adult humans have the ability to digest lactose, the major source of carbohydrate in
milk without any physiological discomfort while some do not. A genetic polymorphism is
responsible for determining the phenotype where some humans express lactase (an enzyme
which helps in digestion of lactose from milk) at high levels throughout their lives, and as a
result are lactose tolerant. While others lose lactase gene expression during childhood, thus
becoming lactose intolerant. Harvey and colleagues (Harvey et al., 1998) reported that 7 polymorphisms in the lactase gene are highly associated and organized in 3 common haplotypes (A, B and C) in individuals of European decent. Thus, the lactose tolerant phenotype evolved from a polymorphic genotype, whereas the lactose intolerant phenotype is a normal phenotype (Ingram, Mulcare, Itan, Thomas & Swallow, 2009).

1.4.2 Transcriptomics

The transcriptome is the complete set of mRNA transcripts that is present in a cell/tissue, or an organism. Transcriptomics is the study of the transcriptome or mRNA transcripts expressed in an organism under a particular condition. Essential and non-essential food components present in the diet can regulate gene expression patterns, which in turn can alter biological responses such as metabolism, cell growth, and cell differentiation. In the context of nutrigenomics, transcriptomics aims at studying the gene expression patterns induced by dietary intervention (a) to elucidate the underlying molecular pathways, such as metabolic pathways and (b) to identify a gene or group of genes that are regulated under healthy or disease conditions (Corthesy-Theulaz et al., 2005; Elliott, 2008; Trujillo, Davis & Milner, 2006).

In the literature, transcriptomic studies have been performed to analyze gene expression changes that take place in the complete absence of nutrients e.g. during fasting state (Bouwens, Afman & Muller, 2007), in the presence of a specific nutrient (van Erk, Blom, van Ommen & Hendriks, 2006), in health as well as in disease conditions (Zemel, Sun, Sobhani & Wilson, 2010).

The most widely used tool for transcriptomics approach is DNA microarray, which allows the measurement of tens of thousands of genes simultaneously (for details refer to Section 2). However, there are a number of challenges and potential pitfalls in using microarray technology (Garosi, De Filippo, van Erk, Rocca-Serra, Sansone & Elliott, 2005):

1. Compromises have to be made in the number and type of samples analyzed due to the high cost of microarrays and to complex logistical issues associated with performing nutritional microarray studies.
2. Unlike the response to drug, the response to nutrients may involve subtle changes in gene expression that may be important in biological terms. However, it may be difficult to detect these changes reliably to differentiate them from the technical noise.
3. Technical variations between array platforms and analytical procedures could lead to differences in the transcriptional responses observed. Consequently, important effects could be missed and/or false leads could be generated.

1.4.3 Proteomics

The term proteome refers to the entire set of proteins present in a cell/tissue, or an organism. Proteomics is the study of proteome under a particular condition (e.g. dietary challenge), where the biological interest is in terms of protein expression, structure, and function. In other words, a proteomics study tries to characterize all proteins in a biological sample (e.g., blood, biopsies tissue) in terms of their relative abundance, distribution, post-translational modifications, functions, and interactions with other biological molecules (Kussmann & Affolter, 2006).

In comparison to transcriptomics relatively less proteomic studies have been reported in the literature. Most of the studies have been carried out using rodent models or human cells in culture (Fuchs, Winkelmann, Johnson, Mariman, Wenzel & Daniel, 2005).

The current technology involves the separation and quantification of proteins by two dimensional gel electrophoresis followed by the identification of proteins by specialized mass spectrometry techniques (fairweather-tait, 2003; Fuchs et al., 2005; Kussmann & Affolter, 2006). Technically, proteomics is very challenging (Kussmann et al., 2006):

1. The human proteome is estimated to encompass several hundred thousand proteins. In addition, different protein forms and variants can exist due to gene and protein splicing, post-translational modifications etc.
2. A single cell may express different sets of protein at different times or under different conditions.
3. The current protein extraction procedures are not always efficient. More specifically, the dynamic range of individual proteins spans 6 orders of magnitude in cells and at least 10 orders of magnitude in the human body (e.g., plasma). At present, mass spectrometry-based protein identification methods cover a dynamic range of 1000 in a single spectrum.
4. The relation between protein structure and function has not been completely understood at the systemic level.
1.4.4 Metabolomics

The term metabolome refers to the complete set of metabolites synthesized/produced by a cell/tissue, or an organism under a particular condition (Corthésy-Theulaz et al., 2005, Harland, 2005). Metabolites are small molecules generated during the process of metabolism. Metabolomics, one of the most recent omics technologies, is the study of the metabolome or global analysis of metabolites and has the goal to provide a comprehensive snapshot of dynamic metabolic pathways (German et al., 2006; Whitfield, German & Noble, 2004).

Biologically relevant samples for metabolomic studies can easily be obtained from blood, urine, saliva, and feces (Kussmann, Affolter, Nagy, Holst & Fay, 2007). Until now, only few examples of metabolomics studies have been reported in human subjects. Most examples have involved the metabolic profiling of individuals, where large scale analyses of body fluids have been used to diagnose metabolic disorders or exposure to xenobiotics (Whitfield, German & Noble, 2004).

There is presently no single technology that can simultaneously identify and quantify all metabolites in a biological sample. Currently, nuclear magnetic resonance spectroscopy and mass spectrometry techniques are being employed for metabolic profiling (German et al., 2006). Metabolomics continues to face challenges such as, sample preparation, technological sensitivity, lack of standardized statistical methods and public databases.

1.4.5 Systems biology

Food is complex and interacts with biological system at different molecular levels. While transcriptomics, proteomics, and metabolomics study these interactions at individual molecular levels, nutritional systems biology integrates the information present at the different levels of genomic expression to obtain a comprehensive understanding of how food affects the human organism as a whole. It is envisioned that the global knowledge of the biological system will improve our understanding of the influence of food on health and disease progression (van Ommen & Stierum, 2002).

Currently, few examples exist in which an integrated approach has been used to examine the influence of exogenous factors on metabolism (Pool-Zobel et al., 2005). A very recent nutritional intervention study was carried out to evaluate the anti-inflammatory properties of certain dietary components on low grade inflammation by analysing the gene, protein and metabolite profiles using different tissues samples (Bakker et al., 2010). This study provides
the evidence for a systemic analysis of the action of nutrients on complex organisms. The main challenge remains to manage and to meaningfully integrate the information obtained at the different omics levels.

1.4.6 Bioinformatics

Each of the omics technologies usually produces/generates high dimensional multivariate data sets. For instance, in microarray-based transcriptomics, a single sample gene expression dataset usually contain expression/intensity values of about 50'000 transcripts. Bioinformatics is therefore a key tool to capture and process the high-dimensional data to extract biologically relevant information. More explicitly, bioinformatics involves the use of powerful computing and software resources for

1. Data acquisition, analysis, and mining using statistical, signal processing, and pattern recognition methods.

2. Creation and management of different databases which also includes setting up of special ontologies that could help in using available resources. Example of these publically available ontologies are KEGG (biological pathway database) (Kanehisa & Goto, 2000), Ensembl (functional annotation database for genes and proteins) (http://www.ensembl.org/index.html), Gene Ontology (controlled vocabulary for gene and protein function) (Ashburner et al., 2000), protein databases (e.g. Swiss Prot).

3. Simulation and modelling. For instance, simulation of complex interactions between genome, nutrition, and health.

One of the main challenges in data analysis and mining is the handling of small biological sample size (Ein-Dor et al., 2006). Reasons such as the cost of the experiment or the sample quality may restrict this parameter.

1.5 Benefits from nutrigenomics

An important question is what benefits can be expected from this new scientific discipline.
Nutrigenomics research will increase our understanding and knowledge of phenotype-genotype-diet interactions in a holistic way with an ultimate aim of developing strategies to improve health management and to prevent diseases. In this regard, the identification of biomarkers that describe the changes from the healthy state to pre-disease and disease states is of paramount importance. Nutritional genomics is believed to deliver new biomarkers (Figure 6) based on expression profiles/patterns generated from genomics, transcriptomics, proteomics, and metabolomics studies. These biomarkers will be indicative of homeostasis or health, pre-disease and disease states (van Ommen, Keijer, Heil & Kaput, 2009). In addition an understanding of how food influences these events in the human organism will greatly improve our present knowledge (Muller & Kersten, 2003; Afman & Muller, 2006; van Der, Stroobant & van Der, 2004).
1.6 Summary

Figure 7 summarizes the general idea of nutritional genomics. Briefly, nutritional genomics is a relatively new approach aimed at understanding the interaction between environment (diet), genotype (genome), and phenotype.

**Figure 7** Overview of the nutrigenomics approach in nutrition research as discussed in this section. The processes of ingestion, digestion and absorption of food/diet facilitates the breakdown of food components into smaller components, which can be accessed and utilized by cells. Cells respond to the available nutrients by modulating mRNA expression, protein expression and metabolite production. These nutrient-induced changes can be detected or studied using high-throughput profiling techniques which can provide information on the molecular pattern related to health, predisease and disease states.

This involves the investigation of multitude of nutrient-related interactions at gene, protein, and metabolite levels.

In this thesis, we present blood cell transcriptomics studies using microarray technology. The following section 2 provides a review on the microarray technology and its application in nutritional genomics.
2. MICROARRAY-BASED TRANSCRIPTOMICS

Transcriptomics study cellular gene expression patterns resulting from endogenous and environmental stimuli. The observed differences in gene expression might either be causative or reflective of changes in the phenotype of the cell/organism studied. As described earlier (Section 1), in the context of nutrigenomics the environmental stimulus is the diet. One of the methods to measure global gene expression patterns is “high-throughput” microarray technology. Most of the microarray experiments can be grouped into three broad categories based on their objectives (Golub et al., 1999):

*Class comparison* - The goal is to compare two different conditions/samples (e.g. healthy vs. diseased) and to identify genes that are “differentially” expressed. For instance, comparing the gene expression of subjects who respond to a treatment/dietary intervention to the gene expression pattern of subjects who do not respond, i.e. comparison between responders and non-responders (Mutch et al., 2007).

*Class prediction* - The goal is to determine/predict, applying multigene statistical models, the class of a new sample using expression profiles obtained from class comparison studies. Class prediction studies are particularly useful for problems such as medical therapy selection, diagnostic classification, prognostic prediction. Let us consider the example of comparison between responders and non-responders mentioned in the previous paragraph again. Using the gene expression data from the comparison study, a prediction model could be built/trained. This prediction model can then be used to predict whether a new subject/patient would respond to a treatment/dietary intervention or not based on his/her gene expression pattern.

*Class discovery* – The goal is to identify new classes of genes that describe the experimental set up of the system studied. Any experiment that is designed to establish which groups of genes are co-regulated can be considered as class discovery.

Strategically a transcriptomic study could be classified in two analytical categories (Muller & Kersten, 2003; Ruegg, Tissot, Farmer, & Mariotti, 2008):

1. *Exploratory analysis* - Gene expression analysis using microarray generates a lot of data and leads to the elucidation of expression patterns. Thus, it is a well
acknowledged method for exploratory analysis, which may lead to the generation of new hypotheses.

2. *Hypothesis-driven analysis* - The experiment is designed to answer a hypothesis that was generated based on previous experiments or a priori knowledge.

Section 2 is organized as follows: Section 2.1 briefly introduces the microarray technology and describes in detail different steps involved in a microarray based transcriptomics study. Section 2.2 provides a literature review on nutritional genomic studies using the microarray technology. Microarray-based nutritional genomic studies require access to a tissue from which the influence of diet on genotype can be sensed effectively. In our work, blood cells are the tissue of interest. In that regard, Section 2.3 introduces blood cell transcriptomics. Finally, Section 2.4 gives a brief overview on an emerging technology used for transcriptomics.

### 2.1 Microarray technology

The microarray technology has evolved from the Southern blotting technique, which was originally developed to analyze complex nucleic acid mixtures (Southern, 1975). A DNA microarray is defined as an orderly arrangement of single stranded DNA molecules on a solid support. The main principle behind microarrays is sequence-specific nucleotide base-pairing between complementary strands of nucleic acids (Figure 8).

![Figure 8 Principle of microarray technology is based on complement base-pairing between nucleotides adenine and thymine; guanine and cytosine between two DNA strands. A:-Adenine, C:-cytosine, G:-guanine, T:-thymine.](image-url)

*Figure 8 Principle of microarray technology is based on complement base-pairing between nucleotides adenine and thymine; guanine and cytosine between two DNA strands. A:-Adenine, C:-cytosine, G:-guanine, T:-thymine.*
The single DNA strand on the array is referred to as the ‘probe’ and the labeled DNA strand extracted from the biological sample (tissue of interest) is referred to as the ‘target’. A DNA microarray can contain tens of thousands of probes.

As described later in detail, the sequence specific nucleotide base-pairing is achieved by hybridization.

Based on the fabrication method and the kind of probes employed, DNA microarrays can be categorized as:

1. **cDNA microarrays** - These microarrays contain immobilized cDNA (complementary DNA) probes or PCR products. The length of the probes typically ranges from one hundred to several hundred nucleotide (nt) bases. One of the main drawbacks of cDNA microarray is that the use of long probe sequences results in cross-hybridization or unspecific hybridization, which in turn leads to the increase of noise level in the data. Currently, cDNA microarrays are less used and have been mainly restricted to gene expression profiling (Yang & Speed, 2002).

2. **Oligonucleotide microarrays** – These microarrays contain synthetic oligonucleotide probes, which are typically between 15 and 100 nucleotide bases long. As the oligonucleotides are much shorter than cDNAs, they allow more specificity and hence are used for gene expression profiling including, genotyping and re-sequencing applications. Long oligonucleotide probes (about 60 nt) have the advantage of providing a high sensitivity for the detection of low abundance transcripts. However, the base composition of oligonucleotides is likely to influence their performance (Kreil, Russell, & Russell, 2006). Oligonucleotide microarrays are fabricated by immobilization (spotting) of pre-synthesized oligonucleotides or by in situ synthesis.

Currently, there are two main commercial distributors of oligonucleotide microarrays, namely, Agilent and Affymetrix. The main difference between these two commercial array platforms is the fabrication method and the manner in which the probes are distributed on the solid support. Figure 9 illustrates the probe fabrication in agilent and affymetrix arrays respectively.
Figure 9 Design of probes on Agilent and Affymetrix arrays. In Agilent arrays, genes are represented as probes of 60-mer oligonucleotide length. The multiple probes representing a single gene are randomly distributed across the array surface. In Affymetrix chip a single probe is represented by 12 perfect match (PM) and mismatch probes (MM).

Agilent arrays use ink-jet spotting methodology to print the pre-synthesized oligonucleotides on the solid surface i.e., glass slides. The oligonucleotides are 60-mer oligo length (i.e., 60 nt). This length gives both specificity and sensitivity to the hybridization of the targets (Hughes et al., 2001).

Affymetrix arrays use photolithography to synthesize the oligonucleotides on the array surface. In contrast to Agilent arrays, the length of the oligonucleotides is 25-mer (i.e., 25 nt). Also, the design of the array is slightly different to facilitate detection of noise. More specifically, the probes are placed such that each gene sequence is represented by 12 perfect match (PM) and mismatch (MM) probes.

Figure 10 presents an overview of the different steps involved in a microarray based transcriptomics study. Broadly, these steps can be grouped into two parts, microarray experiment and microarray data analysis.
2.1.1 Microarray Experiment

The microarray experiment part consists of a) isolation of total RNA from the biological samples of interest to create the target material, b) labeling of the target for microarray detection by using fluorescence (e.g., Cy3 or Cy5, biotin) or chemiluminescence, c) amplification of the targets to increase overall signal strength, d) hybridization of the labeled target to the gene-specific DNA probes on the microarray under conditions that promote only sequence specific annealing, and finally e) measuring the amount of hybridized target on each probe element after removal of un-hybridized target molecules by washing. For fluorescently-labeled targets, a laser scanner is used to generate high-resolution fluorescent images. For the detection of the chemiluminescent labeled targets, high quality array images are obtained by using charged coupled device camera systems. The rest of this subsection describes these steps in detail.
2.1.1.1 Biological sample collection

Gene expression studies usually involve target cells or target tissues from which RNA is extracted. These target cells or tissues can be collected from a tissue culture or from an organism. Due care has to be taken to preserve RNA integrity and abundance during sample collection. If not, RNA molecule may begin to differentiate from their physiological configuration. For example, significant difference in gene expression can occur from ischemia when surgically extracted tissues are kept unprocessed or when RNA is extracted after different times or with different procedures (Huang et al., 2001; Lin et al., 2006; Wittwer et al., 2011). In particular, degradation of RNA by RNases can cause experimental biases (Thompson, Pine, Rosenzweig, Turpaz, & Retief, 2007). One way to preserve RNA integrity and abundance is to lyze and stabilize the samples in solutions that contain RNase inhibitors or that stabilize RNA, such as RNAlater (Ambion). A second option is to snap-freeze the samples in liquid nitrogen at the time of collection. In our work, we used the PAXgene system and RNAlater solutions to preserve the blood samples immediately after phlebotomy. The PAXgene system from PreAnalytiX contains a proprietary reagent that stabilizes RNA for days at ambient temperature. This procedure eliminates the need to process samples immediately after phlebotomy and prevents ex-vivo induction of gene expression (Rainen et al., 2002; Thach et al., 2003).

2.1.1.2 RNA quality assessment

RNA isolation is followed by an evaluation of the quality and integrity of RNA. The quality is usually assessed spectrophotometrically (e.g. ND-1000). Good quality RNA typically has a ratio for the absorbance measured at 260 and 280 nm (A_{260nm}/A_{280nm}) within the range 1.8 - 2.2. A ratio lower than 1.8 may indicate that the RNA is contaminated with protein or other contaminants absorbing at 280 nm. A ratio higher than 2.2 may indicate the presence of contaminants such as, salts and other components in the buffers used to solubilize the RNA. Evaluation of the integrity of RNA is usually conducted with microfluidic capillary systems such as the Agilent 2100 Bioanalyzer or on agarose gels. The Bioanalyzer calculates an RNA Integrity Number (RIN) based on 18S and 28S ribosomal RNA (rRNA) peaks (Schroeder et al., 2006). For microarray analysis, a RIN value greater than 8.0 is considered necessary for high quality samples. Ideally, RNA should be re-purified if the quality criteria are not satisfied. Otherwise, a rule of thumb is that all samples should have “reasonable” and comparable RIN values in order to proceed to the next experimental step. Due to the presence
of high amounts of RNAs in blood, high RIN values for RNA purified from this tissue are difficult to obtain. As such the words ‘comparable’ and ‘reasonable’ take their full meaning. In our experiments RIN values for the RNA ranged between 7.9-9 in the human study and 6.9-8.7 in the mice study.

2.1.1.3 RNA amplification

Each experiment should start with sufficient material so that the mRNA can be directly labeled and measured on microarrays (Scherer, 2009). Usually, a cell contains approximately 20 pg of total RNA, of which <5% is mRNA. In a microarray experiment 5-10 ug of mRNA are typically needed, which may not be obtained from the available samples. Therefore, RNA amplification methods that faithfully reproduce the relative abundance of mRNAs originally present in the samples of interest are employed. There are two main approaches for RNA amplification, namely the linear amplification method, which is based on in vitro transcription, and the exponential amplification method, which is based on PCR (Scherer, 2009).

The linear amplification approach is based on the Eberwine process (in vitro transcription) and is commonly used for DNA microarrays. The amplification process begins with first strand synthesis of cDNA from mRNA by priming with modified oligo(dT) or random primers and by extending the primed oligonucleotides with reverse transcriptase. The primer incorporates a T7 promoter or another bacteriophage promoter onto the 5’ end. After the first round synthesis, a second strand synthesis generates double-stranded DNA by extending the first round cDNA with DNA polymerase. T7 polymerase is then used to synthesize antisense RNA (aRNA). This process can be repeated using aRNA as a template for the second round of amplification (Van Gelder et al., 1990; Scherer, 2009).

Of note, biases may be introduced during the course of amplification due to differences in length of the RNA sequences or in their secondary structure. Furthermore, variations could also be introduced due to differences in incubation time, temperature, or enzymes used in the reactions are modified (Ma et al., 2006).
2.1.1.4 Target labeling

In order to detect cDNA or RNA bound to an array probe these molecules need to be labeled. The labeling molecule can be a fluorescent molecule such as, cyanine (Cy) dyes (Cy3 and Cy5), or biotin. The choice of the labeling strategy depends on the array platform used. The labeling molecules are either introduced directly during reverse transcription (*direct labeling*) or attached after cDNA synthesis by a diesterification reaction using aminoallyl-labeled nucleotides (*indirect labeling*). Indirect labeling is more time consuming but produces higher amounts of labeled molecules and causes less dye effect. It should be noted that differences in labeling efficiency may introduce variation in the measured signal (Russell et al., 2008; Scherer, 2009).

2.1.1.5 Hybridization

The hybridization process involves addition of labeled targets to the array at a specific temperature to allow complementary sequences to anneal. To do so the target first needs to be melted to produce single-stranded oligonucleotides. The main variables that affect the melting temperature and, consequently, influence the hybridization of the target to the probe are temperature, pH, concentration of monovalent cations, and the presence of organic solvents. These variables are optimized to produce signals that, compared to the background noise level, are specific and of high intensity.

High abundance transcripts hybridize with their probes quickly, whereas low abundance sequences take longer. Thus, the reaction time is another critical variable. Thus hybridization of probe and target pairs needs to reach equilibrium so that the true level of each RNA in the sample can be measured accurately (Dorris et al., 2003).

2.1.1.6 Scanning

Microarray scanners are used to measure the amount of each labeled target hybridized to their corresponding probe on the array. There are two types of scanners, namely laser-based scanners and charge-coupled device based imaging systems. The commonly-used laser-based scanner scans the microarray while exciting the fluorophore present on the sample molecules and recording the fluorescence emission with a photomultiplier tube. The fluorescence is captured in the form of a digital image, usually separate 16 bit TIFF (tagged image file format) files for each wavelength scanned. The measured fluorescent signal is assumed to be
directly proportional to the strength of target material hybridized to the probes on the array. The scanned images then become the input for microarray data analysis part (Russell et al., 2008; Scherer, 2009).

### 2.1.2 Microarray data analysis

In microarray based transcriptomic studies, one of the key issues is the reliable retrieval of biologically relevant information from the enormous amount of data that is generated. Since the field is relatively recent, new methods emerge continuously to meet the demands of life science researchers. In that regard, the microarray data analysis consists of 5 part.

a) Obtaining a quantitative readout of array spots. More specifically, array images are processed with array-image processing softwares that convert the brightness of array spots into an intensity value and that associate the spot with gene identity. The gene-specific intensity values from the different samples form the raw data for gene expression analysis.

b) Normalization of the raw data in order to minimize variations induced by the experimental steps.

c) Comparison of gene expression patterns (normalized data) of two experimental conditions using a statistical method such as t-test, ANOVA, or Bayesian statistics to find genes that are differentially expressed.

d) Visualization of the data to identify patterns by clustering techniques or principal component analysis.

e) Inferring biological interpretations from genes that show differential variation in expression between the experimental conditions under study.

The rest of this subsection provides a brief description of each of these processes. For more details, the reader is referred to review paper (Quackenbush, 2007).

#### 2.1.2.1 Image analysis and quality assessment

The process of image analysis extracts fluorescence intensity associated with each arrayed spot from the scanned 16 bit TIFF image (Russell et al., 2008). This process involves:

1. Determination of preliminary spot locations by grid placement.
2. Segmentation of spots into foreground regions (regions with probes) and background regions (regions without probes) using the grid information. Different segmentation algorithms are available to determine the foreground and background regions (Lehmussola, Ruusuvuori, & Yli-Harja, 2006).

3. Extraction of pixel intensities from the foreground and background regions.

In case of Affymetrix and Agilent microarrays, image analysis is performed automatically using feature extraction software. The software also provides different measurements, including mean and median of foreground and background intensities, standard deviation, spot diameter, and pixel regression ratio. Most of these parameters can be used for quality assessment. Diagnostic plots (e.g., scatter plot, box plot, density plot, ratio-intensity plot) generated using these parameters can reveal systematic trends and artifacts.

Finally, spots with very weak signals, high background or uneven signals can be filtered out from the data. The extracted pixel intensities are then used for statistical analysis and interpretation of the data.

### 2.1.2.2 Normalization

As mentioned earlier, the underlying hypothesis in microarray analysis is that the measured intensities for each arrayed gene (feature/probe) represent its relative expression level. In order to identify the biologically relevant patterns of expression, a gene-by-gene comparison is usually performed between the measured expression levels (i.e., intensities) of different states (e.g., before and after treatment).

Let the intensity for gene \( j \) in state \( R \) be \( R(j) \) and in state \( S \) be \( S(j) \), then the gene-by-gene comparison typically involves computation of the ratio:

\[
\text{Ratio} = \frac{R(j)}{S(j)}
\]

A ratio greater than 1.0 means that the gene is up regulated (increase in expression) and a ratio lower than 1.0 means that the gene is down regulated (decrease in expression). However, this simple ratio treats the up regulated gene and down regulated gene differently.
For instance, if the gene expression is up regulated by a factor of 2, then the ratio yields a value of 2, whereas if the gene expression is down regulated by a factor of 2 then the ratio yields a value of 0.5. In practice a logarithm transformation (usually with base 2) is therefore applied on the ratio:

$$\log_2 \frac{R(j)}{S(j)} = \log_2(R(j)) - \log_2(S(j))$$

The logarithm operation treats the up regulated and down regulated genes symmetrically and therefore yields a continuous spectrum of values. For instance, if the gene expression is up regulated by a factor of 2 then the log2 ratio yields a value of +1, whereas, if the gene expression is down regulated by a factor of 2 then the log2 ratio yields a value of -1. After computing the log2 ratio for all genes, the differentially expressed genes are detected using statistical analysis. However, the measured gene intensities may be systematically biased by variations in efficiency of RNA extraction, in DNA amplification (Boelens et al., 2007; Ma et al., 2006), and in labeling efficiency (Lynch, deSilva, Peeva, & Swanson, 2006), as well as by non-uniform hybridization (Schaupp, Jiang, Myers, & Wilson, 2005), scanner instability and variable settings (Shi et al., 2005). Quality checks are therefore important to avoid erroneous findings (Pavlidis & Noble, 2001).

The goal of the normalization process is to remove (or reduce) the systematic variations that affect the measured gene expression levels (Quackenbush, 2002). Different normalization methods have been published, such as global mean normalization, local mean normalization, and quantile normalization. Many of these normalization methods assume that a majority of the genes are not differentially expressed or that the number of up regulated genes roughly equals the number of down regulated genes. We discuss below a few of the normalization techniques that have been used in this work.

**a) Global mean normalization:** Some sources of variation such as differential label incorporation, differing amounts of probe used, or differences in detection efficiency may have a non-uniform effect on the measured gene intensities. In order to handle such variation, a global normalization is performed by dividing all feature intensities by a constant (correction) factor. Alternately, when the logarithm scale is used, the logarithm value of the constant factor intensity is subtracted from the logarithm value of the feature intensities. When the constant factor is the mean intensity of all genes
present on the array, the normalization process is referred to as global mean normalization. Global mean normalization is performed on array by array basis so that global mean is constant across all arrays used in an experiment.

\[
\text{Global Mean Normalized Intensity} = \frac{\text{Feature Intensity}}{\text{Global Mean Feature Intensity}}
\]

b) **Local normalization using Standardization and NOmalization of MicroArray Data (SNOMAD)**: SNOMAD is a tool that has been developed to handle artifactual variations that are not constant across the microarray surface (Colantuoni, Henry, Zeger, & Pevsner, 2002). Global mean normalization cannot address such non-uniform variations as it applies the same transformation to all the feature intensities. In SNOMAD, in addition to (optional) global mean normalization three local normalizations are performed:

1. Firstly, local mean normalization is performed to handle artifacts that are spatially systematic across part of the physical surface of the microarray, such as the ones introduced during the robotic printing of the microarray or in the hybridization. This correction normalizes the feature intensities using a smooth mean estimated locally across the microarray.

\[
\text{Local Mean Normalized Feature Intensity} = \frac{\text{Feature Intensity}}{\text{Local Mean Feature Intensity}}
\]

2. Secondly, local mean normalization is performed to correct bias in gene expression ratios, which varies systematically as the feature intensity changes. This normalization estimates the mean feature expression ratio locally across the range of feature intensities and uses this value to normalize individual feature ratios.

\[
\text{Corrected Log (Ratio)} = \text{Residual} = \text{Log (Ratio)} - \text{local Mean Intensity}
\]
3. The variance in the expression ratios of the genes with low expression levels is often very different from the variance of the genes with high expression levels. The final local normalization corrects this phenomena by standardizing the gene expression ratios to standard deviation calculated locally across the range of elements.

\[
\text{Local Zscore} = \frac{\text{Corrected Log (Ratio)}}{\text{Local Standard Deviation}}
\]

The standardized feature expression ratios can then be used to detect differentially expressed genes. The SNOMAD R script available at http://pevsnerlab.kennedykrieger.org/snomad.php was used in this thesis for the human study.

e) Quantile normalization: The general idea of quantile normalization is to make intensity distributions in each array/slide identical. In other words, after normalization if the intensities on each array are ranked then at each rank position one will find same intensity value across the arrays. The assumption here is that if the same amount of labelled RNA is obtained from two samples and applied to theoretically identical arrays they should have the same overall intensity distribution (Russell et al., 2008). In quantile normalization:

Step 1: Global mean or median normalization is performed on each array. This is also referred to as scaling.

Step 2: The distribution of feature intensities on each array is then centered by mean and variance normalization. This is done by estimating the feature intensity mean (\(\mu\)) and standard deviation (\(\sigma\)), and

\[
\text{Standardized feature intensity} = \frac{(\text{feature intensity} - \mu)}{\sigma}
\]

As a result, on each array the feature intensity mean and variance is 0.0 and 1.0, respectively.

Step 3: On each array, the standardized feature intensities are ranked.
Step 4: For each of the "rank position", the mean of the standardized feature intensities is calculated across the arrays. This results in an average standardized intensity value for each rank position. Note that this average intensity is specific to the rank and not to a feature.

Step 5: Finally, given a feature's rank on the array, the feature is assigned the average standardized intensity corresponding to the rank (estimated in Step 4).

Quantile normalization is integral part of robust multichip average (RMA) normalization that is used for Affymetrix GeneChip (Bolstad et al., 2003; Irizarry et al., 2003). RMA normalization yields normalized feature intensities. The log2 ratio comparing gene-to-gene can then be estimated between two conditions R and S.

2.1.2.3 Detecting differentially expressed genes

Differentially expressed genes could be detected simply by thresholding the log2 ratio. For instance, using a fold change (FC) criteria, the genes with log2 ratio equal to - or greater than a value of +1 (twofold increase) or the genes with a log2 ratio equal to – or smaller than - a value of -1 (twofold decrease) can be considered as differentially expressed. In the early days, microarray studies were using FC to detect differentially expressed genes. However, a criteria such as FC, although simple, has certain drawbacks (Russell et al., 2008). The choice of FC can be arbitrary, and also it may be inappropriate as the “signal-to-noise ratio” (e.g., foreground intensity vs background intensity) is not appropriate for genes with low expression levels as often the case in nutrigenomics studies. Also, in the presence of biological and experimental variability, the FC criteria does not consider gene to gene variability in the significance of the differential expression. Furthermore, FC does not provide information on error rates (i.e., false positive and false negative differential expression of genes). Thus, current studies rely on more robust statistical tests (e.g., hypothesis testing techniques) to identify differentially expressed genes.

One widespread statistical approach is pairwise comparison with the Student's t test to identify differentially expressed genes between two groups – for example, treated and untreated samples. The t-test does not select genes with varying intensities within a condition. However, as some genes can have small standard deviations within a condition and thus be detected as differentially expressed, even if they are not. Modified t-tests, such as Statistical
Analysis of Microarray (SAM), have thus been proposed for microarray (Tusher, Tibshirani, & Chu, 2001).

For experimental designs that have more than two groups of samples, analysis of variance (ANOVA) techniques are usually employed (Churchill, 2004). For instance, in our kinetic experiment in the first study (Chapter 2) we compare untreated sample (i.e., without ingestion of milk or yogurt) against treated sample (i.e., after ingestion of milk or yogurt) at different time points. In ANOVA, a model that takes into account the sources of variation that affect measurements is built, and then the data is used to estimate the variance of each individual variable in the model. ANOVA helps in distinguishing interesting variations, such as gene regulation, and side effects, such as variations caused by different dyes or arrays (Draghici, 2002).

t-test and ANOVA are classical hypothesis testing approaches. In the literature, hypothesis testing approaches based on Bayesian inference have also been proposed for identifying differentially expressed genes (Hatfield, Hung, & Baldi, 2003; Long et al., 2001). In addition, nonparametric methods, such as rank difference analysis method of microarray have also been published (Martin, Demougin, Hall, & Bellis, 2004). During the first phase of the thesis, a rank difference analysis was used to identify differentially expressed genes in the postprandial response of human volunteers to the ingestion of dairy products. This statistical strategy was however not successful in identifying a set of biologically meaningful differentially expressed genes. The reason for this negative outcome has not be elucidated but may be due to a reduce statistical power of the study resulting from the loss of information on the relative expression intensity of each probe following the ranking (F. Schütz, personal communication).

The statistical analyses yield a significance value (also referred to as p-value) for each gene. In other words, the p-value indicates the confidence interval in which a gene can be considered differentially expressed. Due to multiple hypotheses testing, selecting differentially expressed genes solely based on p-value (e.g. thresholding at 5%) can lead to many false positives (i.e., selecting genes that are not truly differentially expressed). Thus, p-values are usually adjusted to estimate the false discovery rate (FDR) for each gene (Benjamini & Hochberg, 1995; Reiner, Yekutiel, & Benjamini, 2003). Differentially expressed genes are then selected using a FDR threshold.
Data visualization for identifying patterns in microarray data

Microarray study usually generates a high dimensional data (expression patterns of about 40'000 – 50'000 transcripts). Even after having identified the differentially expressed genes, one may be left with hundreds or thousands of genes. Therefore, there is need for methods/tools that can assist in organizing and presenting the high dimensional data in a comprehensible manner.

Clustering methods can help to find groups of genes with similar expression profiles across different conditions or samples in the experiment (Eisen, Spellman, Brown, & Botstein, 1998). Different clustering methods, such as hierarchical clustering, K-means clustering, and self organizing map, have been published to investigate microarray data (Russell et al., 2008).

In this thesis, an agglomerative approach for hierarchical clustering has been used (Eisen et al., 1998). The agglomerative approach produces a dendrogram with a bottom-up structure. Firstly, all (individual) data points are treated as separate clusters. The dendrogram is then formed by subsequently merging the two closest clusters based on a distance (similarity or dissimilarity) measured to form new clusters. The distance between clusters can be measured by using the minimum, maximum or the average distance between samples in two different clusters. Figure 11 illustrates the agglomerative approach for hierarchical clustering.

![Figure 11 Clustering based on similarity matrix and average distance measurement.](image)

In addition to clustering methods, dimensionality reduction techniques, such as principal component analysis (PCA) and multidimensional scaling, can be used to visualize the data in a low (two or three) dimensional space. PCA projects the data along dimensions that capture the maximum variance (Raychaudhuri, Stuart, & Altman, 2000). In contrast,
multidimensional scaling projects each data point onto a low dimensional space such that the proximity between the data points in the high dimensional space is maintained in the low dimensional space.

### 2.1.2.5 Biological inference

After statistical analysis of the gene expression data, the biological inference part of analysis can focus along two directions (Figure 12):

a) Finding biological pathways that are enriched by gene expression (pathway enrichment analysis).

b) Finding the regulatory mechanisms that modulate the expression of differentially expressed genes (transcription factor analysis).

![Figure 12 Biological analysis of differentially expressed genes](image)

**Figure 12 Biological analysis of differentially expressed genes**
a) **Pathway enrichment analysis**

A biological pathway could be defined as a series of inter-connected cellular events among biomolecular entities which can be activated by extra-cellular stimuli (Chowbina et al., 2009). The activation of biological pathways leads to changes in the biochemical state of cells.

Using gene expression data from microarrays, pathway enrichment analysis takes advantage of the increasing knowledge on the cellular function of genes to propose biologically meaningful interpretations. More specifically, pathway analysis leads to the interpretation of overall changes in gene expression profiles in terms of functions and processes by integrating gene annotations (e.g., metabolic pathways and gene ontology functional classification) from updated libraries or databases, such as KEGG (Kanehisa & Goto, 2000), Gene Ontology (Ashburner et al., 2000), and Ensemble (http://www.ensembl.org/index.html).

Pathway enrichment analysis can be performed with:

1. A list of differentially expressed genes using threshold criteria based on p-value, FDR, or fold change. Pathway enrichment analysis involves matching/comparing the differentially expressed genes to genes that are part of a specific pathway, and the subsequent identification of matches that are found in larger numbers than expected by chance using statistical tools such as chi-squared tests (Khatri, Bhavsar, Bawa, & Draghici, 2004) and Fisher’s Exact test (Beissbarth & Speed, 2004). The choice of the statistical method usually depends on the number of genes that are enriching the pathway. Another approach is the use of hypergeometric distribution to estimate the probability of finding a specific number of genes belonging to the pathway of interest in the gene list.

2. The entire list/set of genes present on the microarray. In this case, a method referred to as gene set enrichment analysis (GSEA) is typically used (Subramanian et al., 2005). In GSEA, a ranked list of the genes present on the microarray is obtained according to the measure of expression, e.g., fold change or p-value. A competitive statistics (Kolomogrov-Symirnov statistic) is then used to evaluate if extreme parts of the ranked list (i.e., top or bottom) are enriched in the sets of genes that are present in a specific pathway.
Several commercially and publicly available softwares can be used to perform pathway enrichment analysis (Curtis, Oresic, & Vidal-Puig, 2005). In our studies, the software MAPPFinder (Dahlquist, Salomonis, Vranizan, Lawlor, & Conklin, 2002) and GeneGO (MetaCore from GeneGo Inc. USA) were used in combination with differentially expressed gene lists.

b) Transcription factor analysis

The human genome encodes approximately 25,000 genes which constitutes about 1.5% of the genome. The remaining 98.5% of DNA contains complex instructions to trigger gene transcription on or off (Phillips, 2008). Transcription factors are regulatory proteins whose primary goal is to regulate transcription of DNA by binding to promoters sequences (specific consensus sequences) in DNA. Coordinated interaction of multiple proteins at the promoter site forms a transcription complex that facilitates binding and transcription by RNA polymerase which transcribes mRNA. Most eukaryotes have more than 1’500 transcription factors. Regulated transcription begins when a signal is received by a cell (e.g. see Figure 3 section 1 and figure 13). The signal is typically transferred by a signal transduction cascade to a transcription factor. The activated transcription factor enters the nucleus, binds to its promoter binding sites on DNA, and regulates (mostly activates) the expression of specific genes. Through this sequence of events the genes required for a specific tissue at a specific time are expressed appropriately.
As an illustration of the regulatory mechanism, let us consider the following example. NF-κB is one of the major transcription factors which is normally bound to an inhibitor protein and in an inactive state in the cytoplasm (Caamano & Hunter, 2002). An inflammatory stimuli (e.g., induced by the presence of pathogens) triggers a cascade of inter-cellular signalling in the immune cells, which in turn degrades the inhibitor protein bound to NF-κB. The released NF-κB enters the nucleus, binds to promoter regions of genes coding for pro-inflammatory cytokines and induce the transcription, and subsequent translation, of these inflammatory mediators.

The interactions between transcription factors and DNA, control many processes that are important for development and responses to environmental stimuli. Defects in these processes can contribute to the progression of various diseases. In that context, a detailed characterization of the interactions between transcription factors and their promoter regions, as well as of the transcription factor networks can allow a more comprehensive and quantitative mapping of cellular regulatory pathways. In addition, the identification of binding sites for transcription factors on DNA could promote a better understanding of the function of individual genes.
In this thesis, the transcription factor analysis uses promoter sequences in differentially expressed genes to identify transcription factor binding sites. Three different approaches are available (Elnitski, Jin, Farnham, & Jones, 2006):

5. The *pattern matching* approach finds consensus patterns by utilizing prior knowledge on all characterized DNA binding sites for a given protein (transcription factor). This knowledge is available in databases, such as JASPAR and TRANSFAC.

6. The *pattern detection* approach uses de nova pattern detection (i.e., discovery of unknown motifs) to search for recurring and overrepresented patterns in DNA. Typical pattern detection methods are hidden Markov models, Gibbs sampling, and greedy alignment algorithms.

7. The *phylogenetic footprinting* approach identifies potential transcription factor binding sites assuming that the functional sites in promoters should evolve at a much slower rate than DNA regions not carrying conservative functions. This method applies pattern matching approaches on orthologous sequences (i.e., sequences conserved in different organisms) (Dermitzakis & Clark, 2002).

In our first study, we used the CLOVER software tool to find transcription factor binding sites (Frith et al., 2004). For our study, the JASPAR database (Sandelin, Alkema, Engstrom, Wasserman, & Lenhard, 2004) was employed as reference database for pattern matching. The JASPAR database contains a set of “motifs”.

Given the following input information: a set of “promoter sequence” of differentially expressed genes, a reference database and a set of background sequences, in the CLOVER method:

Step 1: The likelihood ratio LR1 of a single motif (from the database) to occur in one particular location of a “promoter sequence” (of differentially expressed gene) is estimated.

Step 2: The likelihood ratio for a motif to be present at any location of the promoter sequence is estimated. This is simply sum of LR1 estimated for different locations on the promoter sequence.
Step 3: Likelihood ratio for combined occurrences (single and multiple occurrences) in a set of promoter sequences is estimated. Let this likelihood ratio be denoted as raw score.

Step 4: Overrepresented motifs are selected based on their p-values which are estimated using background sequences as references. More specifically, each p-value is estimated by

(i) Extracting a random set of sequences from the background sequences matched by length to the target sequences (promoter regions of differentially expressed genes).

(ii) Repeating steps 1-3 to estimate the raw score for the random set of sequences.

(iii) Comparing the resulting raw score with the raw score of the target sequences.

(iv) Repeating steps (i) to (iii) for a large number of randomizations.

(v) Calculating the p-value based on the relative occurrence in which the raw score of the random set of sequences is greater than the raw score of the target sequences.

Figure 14 illustrates the CLOVER steps graphically.
Figure 14 illustration of various steps involved in CLOVER analysis.

2.1.2.6 Software for data analysis

The rapid development of methods for data handling puts high requirements on the software development. Besides a large amount of bioinformatic tools designed to make specific operations, several commercial (e.g. Genespring, ArrayAssist and Kensington Discovery Environment) and open-source (Mev 4, Bioconductor, GenePattern) softwares provide complete solutions (Gentleman et al., 2004; Saeed et al., 2003). Also, the open-source R language for statistical computing has been increasingly popular as a large number of statistical tools are available that can be used to efficiently analyze microarray data.
2.1.3 Microarray data depositories

The massive amount of data generated in a microarray study often exceeds the available analytical resources of the particular research group that produced the data. To accelerate research efficiency, microarray data is now routinely publicly available. Open access to the data was originally primarily realized on the website of the research groups or of the journals publishing the study. Finding and collecting the information from the localized websites was, however, rather time consuming for the researchers.

In response to the exponentially increasing number of microarray data sets published and to urge in conducting meta analyses using several data sets, several public repository data bases for storage of microarray data have been launched. The largest one is the Gene Expression Omnibus at NCBI (http://ncbi.nlm.nih.gov/geo/). In Europe the most popular database is ArrayExpress, which is online since 2002 and which is hosted by the European Bioinformatic Institute (http://www.ebi.ac.uk/arrayexpress/). To streamline the presentation of microarray studies the Microarray Gene Expression Data (MGED) society has proposed a format which holds the Minimum Information About a Microarray Experiment (MIAME) (Brazma et al., 2001). Upon publication of a study it is nowadays often required to have the data publicly available in the MIAME format.

2.2 Application of transcriptomics in nutrition research

In nutrition research, gene expression profiling using microarray technologies can serve three different purposes (Muller & Kersten, 2003). Firstly, it can provide clues on the mechanisms underlying the beneficial or adverse effects of food and nutrients. Secondly, it can lead to the identification of important genes that are altered in pre-disease states and that can be used as molecular biomarkers. Finally, it can assist researchers in identifying and characterizing the basic molecular pathways influenced by food components. Several microarray studies have investigated the molecular effects of foods, food components or nutrition on gene expression in cell cultures, animal models, and humans.

2.2.1 Transcriptomics studies using cell cultures

Cell cultures provide a controlled environment (including the control of growth) to analyse the interaction of dietary components with cellular components. Unlike in vivo models, cells in culture are directly exposed to the dietary components, thus allowing a detailed description
of the molecular pathways influenced by the nutrients. An advantage of in vitro studies is that the experiments can rapidly and economically be conducted and replicated. A limitation of cell culture studies is that the findings may not be directly interpretable in terms of physiological relevance, in particular as these systems are not indicative of the biology of multi-organs environments in vivo (Mortensen et al., 2008). Some of the common cell lines used in nutritional studies are human hepatoma (HepG2), Caco-2, and follicle associated epithelium (FAE). HepG2 can be used to evaluate a range of nutritional factors, in particular metabolic factors. Caco-2 cells may serve as a model system for studying intestinal absorption. FAE can also be used to study the interaction of dietary components with the human gut and their potential to cross the intestinal epithelium (bioavailability). Table 1 summarizes several cell culture studies published in the literature.

Table 1 Examples of transcriptomics studies using cell cultures.

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Dietary component</th>
<th>Findings of study</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caco-2</td>
<td>Iron</td>
<td>Down-regulation of genes involved in cell cycle, DNA metabolism, transcription, protein metabolism, signal transduction, and nucleocytoplasmic transport Up-regulation of genes linked to cell adhesion, lipid and xenobiotic metabolism, iron transport and homeostasis, and immune response</td>
<td>(Bedrine-Ferran et al., 2004)</td>
</tr>
<tr>
<td>Caco-2</td>
<td>Epicatechin (Cocoa)</td>
<td>Prevention of the oxidative DNA damage, reduction of inflammatory response</td>
<td>(Noe et al., 2004)</td>
</tr>
<tr>
<td>HeLa</td>
<td>Zinc</td>
<td>Zinc induced modest changes in gene expression</td>
<td>(Yamada, Suzuki, &amp; Koizumi, 2007)</td>
</tr>
<tr>
<td>THP-1</td>
<td>Zinc</td>
<td>1,045 zinc-responsive genes significantly altered, including cytokine receptors and genes associated with amplification of the Th1 immune response</td>
<td>(Cousins et al., 2003)</td>
</tr>
<tr>
<td>Monocytic cell line</td>
<td>Human milk</td>
<td>Modulation of TLR related signalling</td>
<td>(LeBouder et al., 2006)</td>
</tr>
</tbody>
</table>

2.2.2 Transcriptomics studies using animal models

Animal or human biological systems are more complex than cell cultures. Evidently, cells in culture do not have active communication mechanisms related to multicellular tissue and organ structure, neither a gastrointestinal microflora. Consequently, the full metabolic repertoire of these cells in culture is not expressed to the same level as their complementary in vivo counterparts. Alternately, the metabolism and regulation of nutrient and bioactive components of food are often influenced by the activities of the organs. Animal studies overcome the limitation of cell culture studies by bridging the gap between in vitro systems and whole organisms. Also habitual constraints of human clinical studies, such as the availability of tissues, are removed in animal models. Animal studies could be further
motivated through comparative genomic studies which have demonstrated that animals such as, mice and rats share genes and diseases that are similar to other mammals, in particular man. For instance, 99% of mouse genes have human homologues (Waterston et al., 2002). Similarly, certain diseases such as obesity-induced diabetes occur in humans as well as in mice (Lijnen, 2011; Rossmeisl, Rim, Koza, & Kozak, 2003).

Animal studies typically use laboratory strains, the main advantage being the access of researchers to an array of genetically well defined animal strains resulting from inbreeding. In particular, a huge number of mouse strains have resulted from the efforts to produce and characterize inbred strains for biomedical research (Jackson Laboratory, 2005). Breeding strategies permit the identification of epistatic interactions that are likely to influence gene–disease (Rossmeisl et al., 2003; Cheverud et al., 2004) and nutrient–gene (Cooney, Dave, & Wolff, 2002) interactions. In fact, one of the critical requirements in studies investigating diet-gene interaction in animal models is not the characterization of the selected animal strain but rather the robustness and reproducibility of the molecular composition of the diets investigated (Kaput et al., 2004; Park et al., 1997).

Table 2 shows some transcriptomics studies in animal models. Experimental animals, in most cases mice, have been used to study the effects of caloric restriction, exposure of macronutrients, micronutrients, phytochemicals, and other foods on gene expression in various animal tissues.

Table 2 Examples of animal models used for studying the diet-gene interaction using microarrays.

<table>
<thead>
<tr>
<th>Model</th>
<th>Target tissue/cells</th>
<th>Dietary component</th>
<th>Finding off he study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat</td>
<td>Muscle tissue</td>
<td>High-fat diet</td>
<td>High-fat diet induced reduced expression of genes involved in free-radicle scavenging, tissue development; increased expression of stress response and signal transduction genes (Sreekumar et al., 2002)</td>
</tr>
<tr>
<td>Mouse, Mouse hippocampus</td>
<td>Variation in hippocampal Fe, Cu and Zn concentration</td>
<td>Expression of 25 genes correlated with levels of at least 1 of the 3 metals (Jones et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>Mouse skeletal muscle and heart</td>
<td>Fe overload</td>
<td>Expression of 54 genes and 75 genes affected by iron overload in skeletal muscle and heart, respectively. Genes included involved in glucose and lipid metabolism, gene transcription and cellular stress (Rodriguez et al., 2007)</td>
</tr>
<tr>
<td>Mice</td>
<td>hippocampus and the cortex</td>
<td>Herbal extract (<em>Ginkgo biloba</em>) with neuroprotective properties</td>
<td><em>G. biloba</em> extract affected the expression of genes encoding transcription factors, ion channels, growth factors and neuromodulators, synaptic vesicles and transport, cell surface, and protein kinases and phosphatases (Gohil and Chakraborty, 2004)</td>
</tr>
</tbody>
</table>
2.2.3 Transcriptomics studies in humans

Animal studies do not allow to confirm or refute hypotheses that can be unequivocally translated to human physiology or pathology. However, these models can help generating hypotheses that might be relevant to humans. In short, the findings from cell culture studies or laboratory animal studies must be verified in human subjects. In terms of variations (both due to genetic and environmental factors), human studies are less controllable in comparison to cell culture or animal model studies. Thus, the experimental design plays an important role for an accurate analysis of human studies. For instance, randomized double-blind (where possible) cross-over study designs support an appropriate handling of confounding factors, such as inter-individual variation and bias originating from the subjects or the investigators. In addition, major limitations in human studies are:

1. Target tissue is not easily available. For instance, it is difficult to study the effect of dietary components on liver metabolism or the regulation of appetite in brain tissues.

2. Conducting a long term study is a challenging task as it raises issues related to the life style of the subjects and study compliance.

Unlike cell culture studies and animal studies, only a few nutrigenomics studies have been conducted in human subjects. Table 3 shows a selection of these studies.

Table 3 Examples of studies in human using different tissues.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dietary component</th>
<th>Experimental condition</th>
<th>Findings of study</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adipose tissue</td>
<td>Anti-inflammatory dietary mix</td>
<td>Overweight subjects</td>
<td>Modulation of inflammation, oxidative stress and metabolism</td>
<td>(Bakker et al., 2010)</td>
</tr>
<tr>
<td>Parathyroid tissue explants</td>
<td>High calcium supplementation</td>
<td>Patients with hyperparathyroidism</td>
<td>Modulation of calcium metabolism or calcium related pathways</td>
<td>(Nakajima et al., 2010)</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>Caloric restriction</td>
<td>Obese patients</td>
<td>Down-regulation of CIDEC (cell death–inducing DFF45-like effector) a regulator of adipocyte lipid metabolism</td>
<td>(Magnusson et al., 2008)</td>
</tr>
</tbody>
</table>
2.3 Blood cell transcriptomics in humans

One of the main challenges in human nutrition studies is the tissue availability from healthy subjects. For instance, as mentioned earlier, getting brain tissues or liver tissues is difficult due to ethical concerns. Thus, in analogy to clinical chemistry, nutrigenomic studies often focus on peripheral blood cells as these cells can be obtained in a less invasive manner.

Blood contains a number of different cell types, such as red blood cells, white blood cells, and platelets. Each of these cell types has unique gene transcription profiles which must be taken into consideration when performing transcriptomic studies with blood samples. Although red blood cells (erythrocytes) are by far the most abundant cell type in blood, these cells are transcriptionally inactive as mature erythrocytes are anucleate and, therefore, unable to modify gene transcription in response to environmental changes. Thus, white blood cells are most commonly used for transcriptomic studies in blood samples:

3. White blood cells are exposed to acute and chronic fluctuations in plasma composition arising from the sporadic and varied nature of the diet. These cells therefore react metabolically, also at the level of gene expression, to their exposure to different levels and types of dietary components.

4. White blood cells serve as direct targets for studying the effects of the environment, including the diet, on immune and inflammatory processes.

5. White blood cells act as sentinels, maintaining continuous surveillance for signs of infection or other threats. Thus, they can be used as surrogates for direct sampling of less accessible tissues.

Different types of leukocytes are present in blood. These are generally categorized into neutrophils, eosinophils, basophils, lymphocytes and monocytes. Furthermore, some of these categories can be further sub-divided. The relative proportions of these different cell types can vary somewhat from sample to sample, with knock-on effects on the ‘average’ gene expression profiles (Elliott, 2008).

This noise can be reduced by isolation and analysis of specific leukocyte subsets, i.e. by fractionation of blood cells. Nevertheless, the use of whole blood cells (after globin transcription removal) is advantageous for the following reasons:
1. Multi-centre studies can be conducted as blood can be stored in a stabilizing solution immediately after collection and taken to a different site for processing. In the case of fractionated blood cells, the processing typically needs to be done soon after blood sample collection.

2. Fractionation of blood is avoided, a process that can alter the expression of selected genes in the cells.

3. An average transcript profile of whole blood cells is measured rather than the transcript profile of a specific cell type. Universal metabolic phenomena can thus be identified using the entire population of white blood cell population instead of cell sub-populations with more specific functions.

Another issue encountered when using whole blood samples is the presence of globin transcripts which comprises up to 70% of total mRNA extract. This high abundance of globin transcripts may introduce bias or noise in the data (Wright et al., 2008). To address this issue many experiments have used special protocols to remove the globin transcripts (Field et al., 2007). Other protocols eliminates the need for removing the globin transcripts by having extremely specific DNA–DNA hybridization conditions that strongly minimize unspecific cross-hybridization of the highly abundant globin transcripts to other probes on the array (Parrish et al., 2010). In this thesis, we have employed the first protocols for the human study and the second protocol for the mouse study.

Using blood cell transcriptomics, studies were performed on human subjects to identify factors responsible for variation between individual (inter-individual) and within individual (intra-individual). These studies found that individuals can respond to diet differently. Furthermore, it was observed that intra-individual variation, in the absence of any dietary intervention, is lower than that of inter-individual variation (Whitney et al., 2003; Radich et al., 2004; Eady et al., 2005). This observation led to the suggestion that blood cell transcriptomic studies should be designed in such a way that subjects serve as their own controls, i.e. using a crossover experimental design (van Erk et al., 2006).

Bouwens et al. demonstrated that blood cell can be used to assess the fasting status in humans. Their study showed that, during 24 hours and 48 hours fasting, nutrition-related metabolic changes occur (Bouwens et al., 2009).
Furthermore, a series of blood cell transcriptomic nutritional studies in human subjects have been published since the beginning of this thesis:

4. A low caloric diet down-regulates oxidative stress and inflammatory response in obese men (Crueiras et al., 2008).


6. Fish oil consumption in healthy men has anti-inflammatory effect (Bouwens et al., 2009).


8. Soy isoflavones consumption in healthy postmenauposal women down-regulates cyclin-dependant kinase activity and cell division. The response to isoflavone supplementation depends upon the volunteer’s capacity to produce equol. The production of equol in turn depends on the gut microflora since equol is a bacterial metabolite. Thus, the study presents interesting insight into functional effects of equol in relation to the apparent beneficial effects of isoflavones (Niculescu, Pop, Fischer, & Zeisel, 2007).


10. In comparison with a placebo, anthocyanine supplementation in healthy humans decreased interleukin-8, interferon, and normal T cell expression by 25%, 25%, and 15%, respectively. Inhibition of the key transcription factor NF-κB was identified as the mechanism of action (Karlsen et al., 2007).

11. An anti-inflammatory dietary mix composed of resveratrol, green tea extract, alphatocopherol, vitamin C, n-3 (omega-3) polyunsaturated fatty acids, and tomato extract modulates inflammatory processes, oxidative stress, and metabolism in overweight men (Bakker et al., 2010).
Taken together, these studies clearly show/demonstrates that blood cells can respond to dietary components by inducing specific gene expression profiles. In other words, transcriptomics studies so far have been more observational in nature. Blood cells can thus be a useful non-invasive method to monitor the effects of dietary interventions in human subjects.

2.4 Emerging technology in transcriptomics

Very recently, a technology referred to as “next-generation sequencing” or “RNAseq” has emerged that is changing the way gene expression is studied (Mardis, 2008; Morozova & Marra, 2008). Although applications of this technology to the field of nutrition research have not yet been published, its potential for transcriptomic analyses is underlined by recent applications in other fields of research, in particular for comparative analyses in medicine (Wang, Gerstein, & Snyder, 2009).

The principle behind the application of this novel technique to transcriptomics, has been termed RNA-Seq and is based on a global sequence census approach. Briefly, a complex RNA sample is converted to a library of cDNA with adapters attached to one or both ends. cDNA molecules, with or without amplification (depending on the sequence platform), are then subjected to a massive parallel sequencing of up to 25 million short reads that does not require bacterial cloning prior to sequencing. Following sequencing, powerful bioinformatics tools are needed in order to align the resulting short reads (30-300 bp) to a reference genome or to reference transcripts. The length of these short reads depends on the sequencing platform and the method used. When reference genomes are available shorter reads are sufficient to map their locations rather than sequencing the entire genome. Once mapped, the sequence hits are counted to determine their density and distribution and, consequently, to quantify the gene expression level (Wold & Myers, 2008). RNA-Seq is still a technology under active development that is being evaluated in multiple laboratories for RNA profiling (Torres, Metta, Ottenwalder, & Schlotterer, 2008; Toth et al., 2007). In addition to providing a global information on the level of expression of genes in tissues, RNA-Sep offers the possibility to identify polymorphisms in biological samples. Thus, this technology will soon supplant classical microarray technologies to measure biological transcriptome and to identify polymorphisms, thus accelerating the merging of genomics and genetics in biology and paving the road to personal nutrition research.
2.5 Summary

In this section, we presented microarray-based technology and its application to nutrigenomics.

This thesis presents two nutrigenomic studies using blood cell transcriptomics. In the first study, we investigate the postprandial effect of milk and yogurt on healthy human subjects. While, in the second study we investigate the antimicrobial properties of a potential probiotic strain in mice. Table 4 summarizes the microarray experimental steps and the microarray data analyses steps employed in the two studies.

Table 4 summarizes the microarray experimental steps and the microarray data analyses employed in the two studies.

<table>
<thead>
<tr>
<th>Experimental stage</th>
<th>Human study</th>
<th>Mouse study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Healthy males</td>
<td>C57BL/6J mice</td>
</tr>
<tr>
<td>Target tissue</td>
<td>Whole blood cells</td>
<td>Whole blood cells</td>
</tr>
<tr>
<td>Tissue collection</td>
<td>PAXgene tubes with stabilizing solution</td>
<td>RiboPure tubes with RNAlater solution</td>
</tr>
<tr>
<td>RNA isolation</td>
<td>Total RNA isolation</td>
<td>RiboPure total RNA isolation kit</td>
</tr>
<tr>
<td>Globin transcripts removal</td>
<td>Removed using GlobinClear protocol</td>
<td>None</td>
</tr>
<tr>
<td>RNA amplification</td>
<td>Agilent Linear amplification protocol</td>
<td>SPIA amplification using NutGEN protocol</td>
</tr>
<tr>
<td>Amplified product</td>
<td>cRNA</td>
<td>cDNA</td>
</tr>
<tr>
<td>Target labeling</td>
<td>Direct labeling using Cyanine 5</td>
<td>Direct labeling with biotin</td>
</tr>
<tr>
<td>Hybridization</td>
<td>65°C for 12-16 hours</td>
<td>45°C, overnight</td>
</tr>
<tr>
<td>Scanning</td>
<td>Agilent scanner using CCD</td>
<td>Affymetrix scanner</td>
</tr>
<tr>
<td>Feature extraction</td>
<td>Agilent feature extraction software with automatic gridding</td>
<td>Affymetrix GCOS with automatic gridding</td>
</tr>
<tr>
<td>Normalization</td>
<td>SNOMAD normalization procedure</td>
<td>RMA normalization</td>
</tr>
<tr>
<td>Statistical test</td>
<td>ANOVA and linear contrast analysis</td>
<td>ANOVA and fisher’s least significant difference (LSD) pairwise comparison</td>
</tr>
<tr>
<td>Pathway analysis</td>
<td>GenMAPP, GeneGO software, GSEA</td>
<td>GenMAPP and GeneGO software</td>
</tr>
<tr>
<td>Transcription factor analysis</td>
<td>CLOVER software – Pattern matching algorithm</td>
<td>-</td>
</tr>
</tbody>
</table>

In the context of this thesis, Section 3 gives an overview of milk, its components and their effects on health. Similarly, Section 4 gives an overview on probiotics and its mechanism of action, and presents properties of the potential probiotic strain *Lactobacillus gasseri* K7.
3. MILK AND DAIRY PRODUCTS

Bovine milk and other dairy related products make a significant contribution to human nutrition as well as human health (Haug et al., 2007; Meydani & Ha, 2000). Recently, numerous scientific works have investigated the health benefits of consumption of milk and yogurt that goes beyond the nutritional value. Indeed, dairy products has been associated with reduction of serum cholesterol (Asnaghi, Bruno, Priulla & Nicolin, 2004; Kawase, Hashimoto, Hosoda, Morita & Hosono, 2000), antihypertensive (Jauhiainen & Korpela, 2007; Seppo, Jauhiainen, Poussa & Korpela, 2003), osteoprotective (Moller, Scholz-Ahrens, Roos & Schrezenmeir, 2008), antimicrobial (Haque & Chand., 2008) and immunomodulatory (Madureira, Tavares, Gomes, Pintado & Malcata, 2010; Phelan, Aherne, FitzGerald & O'Brien, 2009; Cross, Stevenson & Gill, 2001; Meydani & Ha, 2000) as well as weight management and obesity prevention or reduction (Zemel, 2005; Zemel & Sun, 2008; Zemel, Sun, Sobhani & Wilson, 2010). Factors responsible for these properties have been attributed to the numerous components present in milk including bioactive components released during digestion or milk processing (Madureira et al., 2010; Moller et al., 2008; Phelan et al., 2009).

In contrast, some research works have also associated milk consumption with negative health effects, such as cardiovascular diseases, obesity, hypercholesterolemia, hypertension, prostate cancer, and metabolic syndrome (Ahn et al., 2007; Melnik, 2009).

Thus, it becomes important to characterize the physiological effects of milk and milk related dairy products, and to tease out the molecular mechanism which may be responsible for the positive and negative effects. Nutrigenomics can help in elucidating the mechanisms by which dairy products and their specific components may influence human health (Casado, Affolter & Kussmann, 2009). In that regard, we can already see efforts in the Milk Genome Consortium that is devoted to the understanding of how various components of milk are synthesized and regulated, and how they achieve their specific functions (Ward & German, 2004).

The remaining part of this section is organized as follows. Section 3.1 describes the different components in milk and their effects on health. Section 3.2 describes yogurt components and their health effects. Section 3.3 describes a few selected metabolic and physiological effects of dairy components relevant in the context of this thesis. Finally, Section 3.4 briefly describes the use of milk as a vector for probiotics.
3.1 Milk components and their health effects

3.1.1 Milk proteins

Bovine milk represents a major source of dietary protein in young and adult humans. The two major protein fractions in bovine milk include micellar casein (α-s1, α-s2, β casein, κ-casein) and soluble whey (β-lactoglobulin, α-lactalbumin, bovine serum albumin, lactoferrin and immunoglobulins) in the ratio of 80:20. Casein micelles can entrap minerals such as Ca\(^{2+}\) and may improve their absorption (Haug et al., 2007).

Milk is rich in essential amino acids and branched chain amino acids (BCAAs). There is evidence that these amino acids, in addition to providing substrates for protein synthesis, also trigger and promote muscle protein synthesis (Yoshizawa, 2004). Not only amino acids but intact proteins in milk, such as β-lactoglobulin, α-lactalbumin, bovine serum albumin, lactoferrin, immunoglobulin and growth factors have been shown to have physiological effects. For instance, the best known immunological properties of β-lactoglobulin are its ability to induce allergic immune responses in susceptible individuals (Knowles and Gill 2004). On the other hand, α-lactalbumin has been shown to increase lymphocytes functions and immune responses of spleen lymphocytes in response to mitogens (Morley et al., 1997; Peterson et al., 1998). While, lactoferrin have been shown to have anti-inflammatory, bacteriocidal properties and antibody response (Knowles & Gill, 2004; Mills et al., 2011). Growth factors identified in bovine milk such as epidermal growth factor (EGF), fibroblast growth factor (FGF1 and FGF2 ), insulin-like growth factor (IGF-I and IGF-II), transforming growth factor (TGF-β1 and TGF-β2) and platelet-derived growth factor (PDGF) have also been shown to regulate growth and development in neonates (Korhonen, 2009; Pouliot & Gauthier, 2006; Playford, Macdonald & Johnson, 2000). Many animal model studies have shown that EGE, IGF-I, and both TGF forms can provoke various local effects on the gastrointestinal tract and can be absorbed intact or partially from intestine into blood circulation and exhibit systemic effect (Gauthier, Yves & Jean-Louis, 2006). The concentrations of all known growth factors in comparison to colostrum are very less in mature milk (Pakkanen & Aalto, 1997). It is noteworthy that the growth factors present in milk seem to withstand relatively well pasteurization and even ultra high temperature (UHT) treatment of milk (Gauthier et al., 2006; Korhonen, 2009).

In addition, milk proteins are currently the main source of biologically active peptides which have been shown to influence various biological functions (Severin & Wenshui, 2005). These
bioactive peptides are either found in the milk as intact protein, released from their precursor protein during gastrointestinal digestion or from the processing of dairy products by lactic acid bacteria (Korhonen & Pihlanto, 2003). Physiological properties of milk derived peptides include antihypertensive, antithrombotic, opioid, mineral binding (casein phosphopeptides (CPPs), antimicrobial, cytomodulatory, and immunomodulatory (Mills et al., 2011; Severin et al., 2005).

Protein fractions present in milk fat globule membrane (MFGM) has also been shown to have therapeutic effect against pathological conditions. For example, fatty acid binding protein (FABP) and glycoprotein in MFGM have been shown to have anti-cancer and antimicrobial properties (Spitsberg, 2005).

One of the negative health effects associated with milk proteins are their role as potential allergens. A susceptible person may be allergic to casein or whey proteins or to both. Symptoms of milk allergy ranges from loose stool, vomiting, fussiness, reduced weight gain to wheezing, and anaphylaxis in extreme cases. Bovine milk allergy may be treated by completely avoiding milk proteins (Wal, 2004).

### 3.1.2 Milk lipids

Structurally, milk fat occurs as micro droplets surrounded by the MFGM, containing mainly triacylglycerols, esters and retinol esters in the core, and primarily phospholipids and cholesterol in the membrane (Argov, Lemay & German, 2008). Milk fat is composed of more than 400 fatty acids with varying length (Mansson, 2008). Broadly, the milk fat can be categorised into saturated fatty acid (65%), monounsaturated fatty acids (25%), and polyunsaturated fatty acids (4.5%). Most of these fatty acids appear in milk as triacylglycerols which make up about 95% of the total lipid/fatty acid. The remaining fat is composed of diglycerides, monoglycerides, free fatty acids, phospholipids, and cholesterol (Haug et al., 2007).

Milk fat has a number of bioactivities and is also important for delivery of fat-soluble vitamins (German & Dillard, 2006). Individual fatty acids have been shown to possess positive health effects. For instance, butyric acid (4:0) is a well-known modulator of gene function, and may also play a role in cancer prevention. Lauric acid (12:0) may have antiviral and antibacterial functions. While, stearic acid (18:0) does not seem to increase serum cholesterol concentration, and is not atherogenic. In contrast to this, the saturated fatty acids
lauric-, myristic-(14:0) and palmitic (16:0) acid have low-density lipoprotein (LDL)- and high-density lipoprotein- (HDL) cholesterol-increasing properties (German & Dillard, 2004). High intake of these acids raise blood cholesterol levels, and diets rich in saturated fat have been regarded to contribute to development of heart diseases, weight gain and obesity (Huang et al., 2007). Because of these negative health effects milk fats especially, saturated fatty acids in human diet have been closely scrutinized in recent years (Melnik, 2009; Mensink, Zock, Kester & Katan, 2003). Despite the high saturated fat content in milk, reviews, on the role of dairy consumption in cardiovascular disease suggest that there is no clear association between intake of dairy products and increased risk of cardiovascular disease. This could be attributed due to the involvement and interaction of multiple factors in milk (van Meijl, Vrolix & Mensink, 2008; German et al., 2009).

Oleic acids is the main mono-unsaturated fatty acids present in milk and is considered to be favourable for health, as diets with high amounts of monounsaturated fatty acid lowers both plasma cholesterol, LDL-cholesterol and triacylglycerol concentrations (Huang et al., 2007). Main polyunsaturated fatty acids in milk include linoleic (n-6 PUFA) and α-linolenic acids (n-3 PUFA). n-6 PUFA have pro-inflammatory properties while n-3 PUFA exhibits anti-inflammatory properties (Russ, Barnett, McNabb, Anderson, Reynolds & Roy, 2010; Haug et al., 2007).

Bovine milk is an important dietary source of the trans-fatty acid, especially conjugated linoleic acid (CLA). CLA has immunomodulatory properties, confers protection against atherosclerosis and diabetes (Bhattacharya, Banu, Rahman, Causey & Fernandes, 2006), has anticarcinogenic activity, and can modulate plasma cholesterol. Milk also contains some vaccenic acid (VA), a trans-fatty acid which has been shown to have adverse health affect by increasing blood lipids. However, more experimental evidence needs to further clarify if VA has unhealthy effects on blood lipids (Haug et al., 2007).

### 3.1.3 Lactose

The most common carbohydrate in milk is lactose, which is specific to mammalian milk. In fermented products such as yogurt, the lactose content is far less (one third) than that of milk because of the conversion of lactose by lactic acid bacteria (LAB) (Kolars, Levitt, Aouji & Savaiano, 1984).
Lactose is hydrolyzed in the small intestine into glucose and galactose by the lactase enzyme found in the intestinal villi. After absorption, galactose is converted to glucose in the liver via Lelior pathway. Deficiency in one or more of the enzymes involved in lactose digestion or galactose metabolism can lead to lactose intolerance or galactocemia, respectively (Schrafsma, 2008). Nutritional properties of lactose include a relatively low sweetening power, calorific value and glycemic index. It also has dietary fibre-like and prebiotic properties, and enhances the absorption of calcium and magnesium in infants but in adults there is no evidence that lactose improves absorption of calcium (Nickel, Martin, Smith, Smith, Miller & Weaver, 1996). Also, slow acid formation by lactose and buffering capacity of milk reduces the cariogenicity of lactose. Furthermore, heat treatment of milk results in the formation of lactulose, a lactose derivative, in small amounts which acts as prebiotics and anti-cariogenic (Schrafsma, 2008).

### 3.1.4 Minerals and vitamins

Dairy products are good source of minerals such as calcium, zinc, magnesium. In particular, calcium content is very high in dairy products. In human nutrition adequate calcium intake is essential. Adequate calcium intake is important for healthy bones and teeth. Furthermore, calcium intake may also help prevent hypertension, decrease the odds of getting colon or breast cancer, improve weight control and reduce the risk of developing kidney stones (Haug et al., 2007). Particular attention has been given to role of calcium in weight management (Zemel et al., 2005).

Milk is a good source of vitamins such as B₁₂, A, and E. These vitamins and selenium have been found to increase the antioxidant properties of milk (Haug et al., 2007).

### 3.1.5 Microvesicles

Recently, it has been discovered that microvesicles (MV$s$) containing milk proteins, mRNA (coding for most milk proteins), and microRNA are present in milk (Hata, Murakami, Nakatani, Yamamoto, Matsuda & Aoki, 2010). Microvesicles released by various cell types, including mammary epithelial cells, may interact with target cells by surface-expressed ligands, transfer surface receptors, deliver proteins, mRNA, and bioactive lipids (Pap, Pallinger, Pasztói & Falus, 2009; Ratajczak et al., 2006; Pap et al., 2009). Ratajczak et al showed that MV$s$ derived from embryonic stem cells may reprogram hematopoietic progenitors by an mRNA-dependent mechanism (Ratajczak et al., 2006). In addition, MV$s$
have been found to show high bioavailability (Golocorbin-Kon et al., 2009). This further suggests a possibility that MVs in milk may act as carriers to transport milk proteins, mRNA and microRNA, lipid bioactive components to target sites and that mRNA present may be involved in cellular reprogramming of the host cell.

### 3.2 Yogurt components and their health effects

Yogurt, in addition to most of the milk components, contains lactic acid bacteria (LAB). Yogurt has been studied for its hypocholesterolemic, immunomodulation, antimicrobial and anticancer effects (Anderson & Gilliland, 1999; Meydani & Ha, 2000). The bacterial and non-bacterial components present in yogurt contribute to beneficial health effects.

Viable LAB in yogurt exhibits antimicrobial properties and this property is attributed to colonization resistance, release of organic acids such as lactic acid and acetic acid, bacteriocines, and other primary metabolites (Lemberg et al., 2007). LAB has also been shown to enhance immune response by regulating the secretion of IgA and certain anti-inflammatory cytokines (Schiffrin, Brassart, Servin, Rochat & Donnet-Hughes, 1997).

Presence of non-bacterial components such as whey protein, short peptides, CLA contributes to yogurt’s beneficial effects. Bacterial fermentation results in more free amino acids, fatty acids and bioactive peptides in yogurt (Chandan and Shahani 1993). The bioactive peptides released during fermentation have been associated with various physiological effects such as antimicrobial, antihypertensive, and immunomodulatory (Paul & Somkuti, 2009).

Finally, in view of various components and their physiological effects, milk can be considered as a complex food matrix and a physiologically rich food. These components in milk may have negative or positive health effects. In fact, the different components may act independently and synergistically with each other to impart their function. With regular discovery of new physiological effects, the role of different components in the milk is getting more and more interweaved (Mills et al., 2011).
3.3 Metabolic and physiological effects of dairy products

When exposed to a changing dietary environment e.g. before/after food intake, cells face a stress condition. They overcome such condition by adaptation i.e., modulating/regulating several cellular signalling cascades. As mentioned earlier in Section 1, the complete cascade of processes by which an extra-cellular signal interacts with a receptor at the cell surface, causing a change in the level of a second messenger ultimately inducing a change in the cell’s function is referred to as signal transduction (Simmons, Fan & Ramabhadran, 2009). Figure 15 illustrates the process of cellular adaptation upon extra-cellular stimuli. Briefly, the extra-cellular stimuli are sensed by transducers (usually kinases and other molecules of signal transduction) which facilitate the activation of transcription factor(s) and initiation of transcription of genes. The transcription of genes first helps the cell to adapt to the new environment and eventually, return to the basal condition.

Milk and dairy products have been reported to have different physiological effects, such as anticancer, antimicrobial, antiinfection, immunomodulatory (anti-inflammatory, antioxidative), antiobesity and hypocholesterolemic effects. In the following subsections, we describe four physiological effects that are central to understanding of chapter 2 and 3. The corresponding signalling mechanism is also presented.
**Figure 15** Mode of activation of the adaptive response of the cell in response to external stimuli. Under basal cellular conditions the transcription factor (TF) is inactive and sequestered or degraded by a “sensor” molecule, usually an inhibitory protein. In the presence of the stimuli, such as dietary components, “transducer” (enzymatic transducers) such as kinases are triggered/activated and modify the sensor, the TF or both. These modifications inhibit the sensor function which induces the liberation, activation and nuclear translocation of the TF, resulting in the up-regulation of key target genes that enable the cell to overcome the stress response. The figure is adapted and modified from (Simmons et al., 2009).

### 3.3.1 Metabolic effect of dairy components

Cells respond to changes in their nutritional/dietary environment quite efficiently and therefore have evolved a host of molecular pathways that can sense nutrients concentration and quickly regulate gene expression and protein modification to respond to any changes. For instance, immune cells constantly monitor the nutrient levels and the energy status to adjust metabolic pathways according to their nutritional status and other environmental stimuli (Lynch, 2001; Peter, Waldmann & Cobbold, 2010). Several nutrient-sensing pathways exist (Lindsley & Rutter, 2004). Among which the mammalian target of rapamycin (mTOR) pathway has emerged as one of the central and key nutrient-sensing pathway. mTOR responds to nutrient and growth factors and mediates cell
growth by positively regulating (i.e., activating) anabolic processes such as translation, the transcription of ribosomal genes and the synthesis of ribosomal proteins and transfer RNA, mitochondrial activity and by negatively regulating (i.e., deactivating) catabolic processes such as protein and RNA degradation (Asnaghi et al., 2004; Blagosklonny & Hall, 2009; Lindsley et al., 2004; Wang & Proud, 2006).

Amino acids are positive regulators of mTOR signalling while the drug rapamycin inhibits its activity (Lynch, 2001). Among the amino acids, leucine is a potent stimulator of the mTOR pathway in mammalian cells (Proud, 2002). However, the basic mechanism behind the amino acid sensing phenomena preceding to mTOR activation (e.g. what actually binds amino acids) still remains unclear (Hall, 2008; Lindsley et al., 2004).

mTOR is a serine/threonine protein kinase composed of two complexes mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2). mTORC1 regulates protein synthesis by acting upon two of its substrates, namely 40S ribosomal protein S6 kinase (p70S6K) and eIF4E-binding proteins (4E-BP1) (Van Der, Beyaert, Inze & De Veylder, 2009). mTOR kinase, in response to amino acids and growth factors, phosphorylates its substrates, inducing activation of p70S6K and inhibition of 4E-BP1(Figure 16). The activated p70S6K in turn phosphorylates the 40S ribosomal protein S6 (RPS6) and possibly the translation initiation factor eIF4B (Hall, 2008). 4E-BP1 is a translation inhibitor which binds and inhibits the translation initiation factor eIF4E. Phosphorylation by mTOR kinase induces its dissociation from eIF4E, which can then bind the cap structure at the 5’ termini of mRNAs, thereby allowing cap-dependent translation (Van Der et al., 2009).

As mentioned earlier, whey protein and casein protein fractions in dairy products are rich in essential amino acids, especially BCAAs. Leucine, one of the BCAA, is a potent activator of muscle protein synthesis through mTOR mediation (Yoshizawa, 2004). mTOR mediated protein synthesis in muscle cells increases muscle mass (Norton & Layman, 2006).

In immune cells, increased protein synthesis process has been associated with increased metabolic activity (Januszkiewicz et al., 2002). In that regard, mTOR links nutrient sensing and the immune response (Hotamisligil & Erbay, 2008).
Figure 16 Amino acids (leucine) and insulin/growth factors stimulate mTOR activity while low energy status of a cell inhibits mTOR activity. Activated mTOR phosphorylates its substrates (p70S6k and 4E-BP1) and stimulate protein synthesis. P70S6k: p70S6 kinase, 4E-BP1: eukaryotic translation initiation factor 4E binding protein 1. Figure adapted and modified from Lynch 2001.

3.3.2 Anti-oxidative and anti-inflammatory effects of dairy components

Ingestion of glucose, protein, lipid and a mixed meal provokes an oxidative and inflammatory response (Aljada et al., 2004; Mohanty, Ghanim, Hamouda, Aljada, Garg & Dandona, 2002; Aljada et al., 2006; Mohanty, Hamouda, Garg, Aljada, Ghanim & Dandona, 2000; Esposito & Giugliano, 2006). Immune cells exposed to acute postprandial levels of nutrients react, first by mounting a transient oxidative and inflammatory response, then by counteracting this response with an adaptive response. This oxidative and inflammatory response coined as “postprandial stress”, is transient in nature and low in intensity compared to that of a classical inflammatory disease (Margioris, 2009) (Figure 17).
**Figure 17 Postprandial stress after macronutrient ingestion as published in Margioris, 2009.**

Oxidative stress is caused by reactive oxygen species (ROS) and reactive nitrogen species (RNS). Oxidative stress is harmful to cells as it causes damage to cellular lipids, proteins, and DNA. Thus, oxidative stress has been implicated in onset and the progress of a number of human diseases, such as obesity, cancer, diabetes mellitus, inflammatory diseases, neurogenerative disorders as well as in the ageing process (Limon-Pacheco & Gonsebatt, 2009). Oxidative stress is a cellular imbalance with, on one side an over production of ROS and on the other side, a deficiency in enzymatic and nonenzymatic anti-oxidants. Living organisms are protected from various oxidative stresses by maintaining “redox homeostasis” through antioxidant enzymes.

NRF2 is one of the key redox-sensitive transcription factor that activates the expression of the most anti-oxidant and cytoprotective genes upon exposure to ROS following its binding to antioxidant response elements (ARE) in DNA (Paul et al., 2009; Kang, Lee & Kim, 2005) (Figure 18).
Figure 18 Pathway describing the mechanism of activation of NRF2 and the induction of antioxidant / cytoprotective gene expression. Under basal condition, NRF2 is inactive in the cytoplasm and bounded to an repressor protein, Keap-1. Increase in ROS levels induces Keap-1 to undergo modification and NRF2 is phosphorylated by PKC. This results in dissociation of NRF2 and Keap-1 complex. Phosphorylated NRF2 translocates to the nucleus and binds the antioxidant response element (ARE) with another transcription factor, MAF. This leads to an antioxidant and cytoprotective gene expression that protects the cell. Alternatively, the NRF2 protein constantly undergoes proteosomal degradation, there by shortening the half-life of NRF2 protein in the cytoplasm (Wright & Dennery, 2009).

Various components in bovine milk and yogurt possess antioxidative properties. Bovine milk has antioxidant components in the form of enzymes such as superoxide dismutase, catalase and selenium-containing glutathione peroxidases (GSHPx), and a non-enzymatic fraction containing lactoferrin, vitamin C and vitamin E (Lindmark-Mansson & Akesson, 2000). In addition, peptides with antioxidative properties can be released from caseins and whey proteins by hydrolysis with digestive enzymes and by proteolytic LAB in fermented milks (Korhonen & Pihlanto, 2003; Mills et al., 2011).

Human milk also has been shown to have antioxidant properties (Ledo et al., 2009). In particular, breast milk provides better antioxidant activity than formula milk and helps in eliminating ROS (Aycicek, Erel, Kocyigit, Selek & Demirkol, 2006).
Constant exposure to meals high in certain nutrients stresses the immune/metabolic balance. As a result, immune cells mount their first line of defense to encounter such imbalance in homeostasis. Toll-like Receptors (TLRs) are key players in mounting the innate immune response during stress. TLRs are a family of pattern-recognition receptors. To date, 12 members of the TLR family have been identified in mammals (Akira, Takeda & Kaisho, 2001). Upon binding to their ligands (e.g. lipopolysaccharides from gram negative bacteria), TLRs recruit adaptor molecules to their intracellular signalling domains leading to the activation of several kinases. This in turn results in the activation and translocation of NF-κB, a redox sensitive transcription factor (like NRF2) that regulates pro-inflammatory genes (figure 19). This leads to the inflammatory response (Akira et al., 2001).

Recent studies indicate that saturated fatty acid (SFA) and polyunsaturated fatty acids (PUFA) are capable of modulating TLR4 signalling with opposite effects. SFA activates TLR4 signalling and induces the activation of NF-kB transcription factor, while PUFA suppresses the activation of TLR 4 signalling. Though the exact mechanism is unclear SFA and PUFA may interfere with the structure of membrane lipid raft (Wong, Kwon, Choi, Kim, Nakahira & Hwang, 2009). Interestingly, there are also studies suggesting that ingestion of diet rich in fat, especially long chain fatty acids, facilitates the absorption of lipopolysaccharides (LPS) present in the gut lumen, which binds to TLR4 and enhances the activation of downstream signalling, including NF-kB activation (Ghoshal, Witta, Zhong, de Villiers & Eckhardt, 2009).
The choice of diet can profoundly modulate or reduce the postprandial stress (oxidative and inflammatory stress). Several food and dietary components can have a beneficial effect in reducing oxidative and inflammatory stress. For example, the intake of orange juice by healthy volunteers, along with a high-fat high-carbohydrate diet, markedly decreases the postprandial oxidative stress and the inflammatory response. The study also showed a significant decrease in expression of the inflammatory markers TLR4 and TLR2. The observed effect has been attributed to the antioxidants present in orange juice (Ghanim et al., 2010).

Dairy product especially milk, have been shown to reduce inflammatory markers in overweight and obese subjects. This anti-inflammatory effect has been attributed to calcium and other milk components such as, BCAA, or ACE-inhibitory peptides (Zemel et al., 2008). In a different study conducted with overweight and obese subjects and comparing a dairy diet
with a soy diet, dairy products exhibited anti-oxidative and anti-inflammatory effects while the soy diet did not (Zemel et al., 2010). Very recent study by Zemel et al reported that dairy products attenuate oxidative and inflammatory stress in the metabolic syndrome. Furthermore, adequate intake (3.5 serving/day) of dairy products in comparison with a low intake (0.5 serving/day) is more effective in inducing these effects (Zemel et al., 2011, unpublished).

In addition, in vitro studies have shown that bioactive components in milk modulate various aspects of immune function (Gill, Doull, Rutherford & Cross, 2000). For instance, CLA from milk exhibits an anti-inflammatory activity by suppressing NF-kB activation (O'Shea, Bassaganya-Riera & Mohede, 2004). Similarly, lactoferrin and TGF-β from milk activates IL-10, an anti-inflammatory cytokine (Donnet-Hughes, Duc, Serrant, Vidal & Schiffrin, 2000; Döring, Will, Amasheh, Clauss, Ahlbrecht & Daniel, 1998).

### 3.3.3 Cytomodulatory effect of dairy components

Apoptosis, or programmed cell death, is a highly regulated process, important for the maintenance of tissue homeostasis and for specific tasks such as the regulation of immune cell selection and activity (Fadeel & Orrenius, 2005). The process of apoptosis helps in elimination of damaged and unwanted cells by allowing cells to undergo self-destruction. Unlike necrosis, another form of cell death, where the cell dies by swelling and bursting its content in the area causing an inflammatory response, apoptosis is a controlled process where the content of the cell is kept strictly within the cell membrane as it is degraded. The apoptotic cells are then phagocytosed before the cell’s contents leak into the neighborhood (Sun, 2008). Therefore, apoptosis can prevent unnecessary inflammatory response.

The mechanism of apoptosis is complex and involving several cascading molecular events. Apoptosis in a cell, can be triggered by two major pathways, (Figure 20) namely, the intrinsic and the extrinsic pathways. The intrinsic pathway involves mitochondrial mediated cell death. Multiple cellular stress signals, including oxidative stress, DNA damage, ROS, induces the pro-apoptotic proteins present in the cytoplasm, BCL2-associated X protein (BAX), BH3 interacting domain death agonist (BID) and BCL2-antagonist/killer (BAK) to bind to the outer membrane of the mitochondria and causes the release of cytochrome c (cyt c) and
other intramembrane apoptogenic proteins from the mitochondria into the cytosol. In the cytosol, cyt c forms a complex with the caspase-activating protein, apoptotic protease-activating factor 1 (Apaf1). Following its formation, the complex activates caspase-9, a initiator protein. The activated caspase-9, together with the complex of cyt c and Apaf1 forms a multiprotein structure known as the “apoptosome.” Activation of apoptosome associated cell death protease caspase-9 then initiates a proteolytic cascade, where activated caspase-9 cleaves and activates downstream effector protein, caspase-3 that initiates degradation. In contrast, the extrinsic apoptotic pathway relies on tumor necrosis factor (TNF) family death receptors for triggering apoptosis. A subgroup of the TNF family receptors contains a cytosolic death domain, which enables their intracellular interaction with downstream adapter proteins, linking these receptors to specific caspses. Upon ligand binding, TNF family receptors containing cytosolic death domains cluster in membranes, recruiting caspase binding adaptor proteins, including the adapter Fas associating protein with death domain (FADD). FADD activates caspase 8 which in turn activates caspase 3 to initiate cell degradation (Reed 2002).

The regulation of apoptotic events in the mitochondria occurs through activity of members of the B-cell CLL/lymphoma protein 2 (BCL2) family of proteins and inhibitors of apoptosis proteins (IAP). Members of the BCL2 family of proteins may be pro- (BCL-10, BAX, BAK, BID, BAD, BIM, BIK, and BLK) or anti-apoptotic (BCL-x, BCL-xL, BCL-xS, BCL-w, and BAG) BCL2 proteins regulate apoptosis by their ability to homodimerize or heterodimerize at the mitochondrial membrane, and it is the differential recruitment of pro- or anti-apoptotic family members that is thought to tilt the balance between cell death and survival. IAP proteins act downstream to inhibit activity of caspase 9 and 3 (Salvesen & Duckett, 2002). Figure 20 illustrates the two apoptotic pathways.
Figure 20 Apoptosis can be initiated by the extrinsic pathway which involves death-receptors. Upon ligand binding, fas associated death domain (FADD) is activated that acts through caspase 8 to activate caspase 3. The intrinsic pathway involves mitochondria. Various stress stimuli including ROS activates the intrinsic pathway by activating pro-apoptotic proteins (BID) in the cytoplasm. Pro-apoptotic proteins bind to the mitochondrial and promote the release of cytochrome c. In the cytoplasm, the released cytochrome c forms complex with apoptotic protease-activating factor 1 (APAF1) and recruit caspase 9 to activate caspase 3. Thus, both pathways converge to activate the caspase 3, which initiates cell degradation. In addition, apoptosis is regulated by the B-cell lymphoma protein 2 (BCL2) and inhibitor of apoptosis (IAP) protein families. BCL2 proteins are thought to regulate the mitochondrial permeability transition by inhibiting (BCL2, BCL-XL and MCL1) or promoting (BAX and BID) cytochrome c release. IAP proteins act downstream to inhibit the activity of caspase 9 and caspase 3. Figure adapted and modified from (Andersen et al., 2005).

Transcription factor NF-κB induces the expression of pro-survival BCL2 members (Karin, 2002), and also positively contribute to apoptosis induction by activating a number of other antiapoptotic genes, such as the IAPs (Zou et al., 2004). Under certain circumstances it can promote apoptosis, though the mechanism is unclear (Grimm, Bauer, Baeuerle & Schulze-
Osthoff, 1996; Karin & Lin, 2002). However, a variety of mechanisms may exist which may modulate the activity of NF-kB and thereby affect the ultimate outcome of a cell's fate (Karin et al., 2002).

A delicate balance between pro-apoptotic and anti-apoptotic regulators of apoptosis is essential for ensuring the survival of long-lived cells and the proper turnover of short-lived cells in a variety of tissues, including the bone marrow, thymus, and peripheral lymphoid tissues. However, imbalances in this delicate balance of pro-apoptotic and anti-apoptotic proteins occur in disease scenarios, including cancer, where the scales tip in favor of antiapoptotic proteins, resulting in cells with a selective survival advantage that promotes neoplasia and malignancy (Reed, 2002).

In literature, it is suggested that dietary components, can influence the process of apoptosis (Watson, Cai & Jones, 2000; Martin, 2006; Martin, 2007). For instance, some phytochemicals (such as resveratrol, a polyphenol in red wine and grapes), Beta carotene (a carotenoid in orange vegetables), antioxidants (such as vitamin E, vitamin C and selenium) selectively induces apoptosis in cancer cells without affecting normal cells (Martin, 2006).

Various components of milk have been proposed to have anticancer properties (Gill & Cross, 2000). It has been proposed that milk fat components such as conjugated linoleic acid, butyric acid and sphingomyelin are potential anticarcinogenic agents (Parodi, 1997). Calcium in milk products has been found to give protection against colon cancer, while in other studies, it has been found to promote prostate cancer (Tsuda et al; 2000). The major milk proteins (caseins, whey and MFGM protein fractions) have also been shown to possess anticancer properties. For example, in colon cancer models in rats, whey protein concentrate was found to be protective against cancer by increasing the intracellular levels of glutathione. In addition, whey fractions such as lactoferrin and α-lactalbumin have been shown as apoptosis-inducing agent (Gill & Cross, 2000; Tsuda et al., 2000). Furthermore, one of the proteins isolated from MFGM termed as fatty acid binding protein (FABP) has been shown to inhibit some breast cancer cell lines (Mather, 2000).
### 3.3.4 Anti-hypertensive effect of dairy components

Hypertension is a risk factor for cardiovascular diseases, including coronary heart disease, peripheral arterial disease, and stroke. The renin-angiotensin system (RAS) is an important regulator of blood pressure (figure 21).

![Renin-angiotensin system](image)

**Figure 21** Renin-angiotensin system (RAS) regulates blood pressure. RAS has two axis. On one axis, rennin cleaves its substrate angiotensinogen into angiotensin I. Angiotensin converting enzyme (ACE), in turn cleaves angiotensin I to angiotensin II which is a vasoconstrictor. On the other axis, angiotensin converting enzyme 2 (ACE2) cleaves single amino acid from angiotensin I and angiotensin II, giving rise to angiotensin (1-9) and angiotensin (1-7) peptide respectively. Angiotensin (1-7) peptide is a vasodilator.

In the classical RAS, the protease renin cleaves angiotensinogen into the inactive decapeptide angiotensin I (AngI: Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu). Angiotensin-converting enzyme (ACE) then catalyses the cleavage of AngI into the active octamer angiotensin II (AngII: Asp-Arg-Val-Tyr-Ile-His-Pro-Phe), which contributes to hypertension by acting as a vasoconstrictor (Beierwaltes, 2010). Recently, a counter regulatory axis of RAS has been identified in which a homologue of ACE, termed ACE2, is the main component. Unlike ACE, ACE2 functions as a carboxypeptidase, cleaving a single amino acid residue from AngI, generating AngI-9, and a single amino acid residue from AngII to generate Ang1-7, a
vasodilator (Figure 21). ACE2 seems to inhibit the formation of angiotensin and prevent vasoconstriction by enhancing vasodilation (Bernstein, 2002; Crackower et al., 2002). In vitro studies suggest that ACE2 may modulate the RAS and thus affect blood pressure regulation. Crackower et al demonstrated that ACE2 (ace2 -/-) knockout mice showed significant heart defects while further deletion of ACE resulted in restored cardiac functions (Crackower et al., 2002).

Dairy products are good source of antihypertensive peptides (Butikofer, Meyer, Sieber, Walther & Wechsler, 2008). Milk-derived bioactive peptides inhibit the activity of the ACE enzyme and, may thus regulate blood pressure (BP). These ACE inhibitors are thought to play a dual role by preventing ACE from synthesizing the potent vasoconstrictor, angiotensin-II as well as preventing the enzymatic degradation of bradykinin, a vasodilator (Butikofer et al., 2008; Seppo et al., 2003).

Milk fermentation using LAB or their proteinases has been proposed as a strategy to release ACE-inhibitory peptides from milk proteins, especially caseins (Hayes et al., 2007; Seppo et al., 2003). For example, the ACE-inhibitory tri-peptides Val-Pro-Pro (Clare, Catignani & Swaisgood, 2003; Juillard, Laan, Kunji, Jeronimus-Stratingh, Bruins & Konings, 1995) and Ile-Pro-Pro (Chabance et al., 1995) derived from casein following fermentation with the strains Lactobacillus helveticus and Saccharomyces cerevisiae, respectively, are one of the well-studied peptides. These bioactive peptides have also been commercialized as sour milk product, called Calpis (Calpis, Co. Ltd., Tokyo).

Administration of Calpis to spontaneously hypertensive rats (Naoyuki Yamamoto, Atsuko Akino & Toshiaki Takano, 994) as well as mildly hypertensive patients (Hata, Yamamoto, Ohni, Nakajima, Nakamura & Takano, 1996) resulted in a significant reduction in blood pressure. More recently, casein hydrolysate containing Val-Pro-Pro and Ile-Pro-Pro has been shown to improve vascular endothelial dysfunction in subjects with mild hypertension (Hirotot et al., 2007). Recent studies, however, challenges the possibility that these peptides can reach pharmacological concentration to efficiently decrease blood pressure (van Mierlo et al., 2009; Engberink et al., 2008).

3.4 Milk as carrier for probiotics

Milk and other dairy products are best suited for lactic acid bacterial growth as milk has a good buffering capacity (Paul et al., 2009). Thus, in addition to providing nutrients and
bioactive components milk and dairy products can serve as carriers for “probiotics”, especially lactic acid bacteria. As bacterial fermentation may influence the release of bioactive components in milk, the addition of probiotics bacteria may give added value to milk and other dairy products.

In summary, milk is a complex food matrix whose components may act in synergistic and antagonistic ways in order to impart health effects. Also, there is growing recognition that nutrients, when present within the food matrix exert more health promoting effects than as isolated supplements (Virgili & Perozzi, 2008). Jacob and Tapsell proposed that identification of bioactive components in food may be helpful in explaining the health effects of food. However, such information is incomplete without understanding the context in which these bioactive agents operate in the physiology of the food source (plant or animal) (Jacobs, Jr. & Tapsell, 2007). It is thus important to study food in its complexity to get a complete picture.

In that context, Chapters 2 and 3 of this thesis present a blood cell transcriptomic study that investigates the effect of ingestion of whole milk and yogurt on healthy human subjects.
4. PROBIOTICS

Probiotics, as defined by World Health Organization, are live micro-organisms which, when administered in adequate amounts, confers a beneficial effect on the health status of the host (Food and Agriculture Organization of the United Nations. Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria., 2011) The concept of probiotics can be traced back to the work of Ellie Metchnikoff who associated longevity of Bulgarian peasants to their dietary intake of fermented milk products. In particular, he suggested that the live microbes in the fermented milk products maintain a good balance of the intestinal microbiota thereby influencing health (Metchnikoff, 1907). Today the probiotic research has gained wide attention in the dairy research community.

Extensive research is usually needed in order to label a microorganism as a probiotic and requires documentation of scientific evidence to substantiate health benefit associated with probiotics. One of the main prerequisite for microorganisms to be classified as probiotics is that they should be safe for human use. The safety assessment of each potential probiotic strain involves a thorough *in vitro* evaluation followed by *in vivo* studies. The efficacy of the probiotics needs to be evaluated by at least one human trial which should be double-blinded and placebo-controlled (Dunne et al., 2001; Saarela, Mogensen, Fondén, Mättö & Mattila-Sandholm, 2000).

*Lactobacillus* species and Bifidobacterium species are popular probiotics choice due to their long history of safe use in the fermentation industry and to the fact that they are natural inhabitants of the gastrointestinal tract (GIT). The most common *Lactobacillus* species isolated from the human intestine includes the *Lactobacillus* acidophilus group of LAB, *L. salivarius, L. casei, L. plantarum, L. fermentum, L. reuteri* and *L. brevis* (Saito, 2004). LAB isolated from human feces or the intestine, are thought to have beneficial effects on health (Tannock, 1997). Other common genera used in probiotic preparation are *Streptococcus, Lactococcus, E. coli*, and some fungal strains (Saxelin, 2008).

Probiotics have the following properties and functions (Isolauri, Juntunen, Rautanen, Sillanaukee & Koivula, 1991; Kaur, Chopra & Saini, 2002):

1. Acid resistance and bile tolerance - The efficiency of the probiotics depends on their ability to survive the GIT by overcoming the low pH in the stomach, tolerate the bile acids and pancreatic juices to reach the target site. One of the mechanisms proposed to
counteract the toxic effect of bile salt existing in the intestinal environment is the presence of a gene which codes for bile salt hydrolase enzymes. The enzymes help in hydrolysis of bile salts to make them less toxic for the lactic acid bacteria.

2. Adherence to host epithelial tissue - Once the probiotic survives the harsh condition of GIT, the next step is its adherence to the epithelial cells in the intestine in order to colonize it. In vitro systems such as Caco-2 cell and HT-29 cell lines serve as a suitable and valuable tool for assessing the adherence properties of potential probiotic strains.

3. Colonization of the intestine - The ability of probiotics to adhere to intestinal cells and to further colonize it is important to impart their beneficial effects on the host. For example, L. rhamnosus GG has a longest retention time in the GIT. As a result, it has the ability to shorten or limit rotavirus-associated diarrhea in infants (Dunne et al., 2001; Isolauri et al., 1991). In contrast, L. bulgaricus and S. thermophilus do not colonize the intestine and are ineffective in rotavirus-induced diarrhea (Boyle, Robins-Browne & Tang, 2006).

4. Inhibition of the pathogenic adherence and elimination of pathogens.

5. Production of acids, hydrogen peroxide, and bacteriocins that antagonise pathogen growth.

6. Improvement of intestinal microflora and intestinal gut health.

7. Safety, i.e. non-pathogenic and non-carcinogenic.

In recent years, many studies have been carried out to establish the health benefits of probiotics. Probiotics have been associated with the prevention of diarrheal disease, cancer, food allergy and upper respiratory tract disease, lactose utilization, stabilization of the gut mucosal barrier and immunomodulation (Kailasapathy & Chin, 2000). Among these, many antimicrobial properties and immunomodulation of the host have received the most attention (Gill & Prasad, 2008).
4.1 Mechanism of action of probiotics

The GIT of an organism has a dynamic environment as it is continuously exposed to a complex and dynamic community of microorganisms, including commensals, probiotics and pathogens. The GIT plays an important role as an interface between the host and the gut lumen environment. It is home to about 100 trillion microbes of many different species (O’Hara & Shanahan, 2006; Xu et al., 2003; Musso, Gambino & Cassader, 2010). Intestinal epithelial cells have the capacity to distinguish pathogenic from nonpathogenic bacteria on the basis of their invasiveness and the presence of flagella, cell wall components, although the exact mechanisms that allow them to distinguish between pathogenic and non-pathogenic bacteria have not been fully elucidated (Borchers, Selmi, Meyers, Keen & Gershwin, 2009). In addition to this, mounting evidence suggests that the immune cells underlying the basolateral side of the gut epithelium, such as macrophages and dendritic cells, are in constant dialogue or cross-talk with intestinal cells and the microflora, including probiotics and even pathogens. This crosstalk between immune cells and the gut microflora plays a significant role in regulating the immune response in the host (Clavel & Haller, 2007; Corthesy, Gaskins & Mercenier, 2007; Galan & Bliska, 1996; Haller, 2008; Hörmannsperger & Haller, 2010). Sufficient evidence suggests that LAB exert their immune enhancing effects by modulating specific (e.g. antibody production, cytokine production, lymphocyte proliferation) and non-specific (e.g. phagocyte function, NK cell activity) immune responses in the host (Gill, 1998; Gill & Prasad, 2008).

Though the precise mechanism of action of probiotics has not been fully elucidated, many potential mechanisms or mode of actions have been proposed. Probiotics could influence the host at three different levels, namely the gut microflora, the intestinal epithelium, and the immune system (Gareau et al., 2011; Michail S, 2005).

4.1.1 Influence of probiotics on the gut microflora

One of the mechanisms of action of probiotics is their ability to influence the intestinal microflora. Probiotics can affect the microflora including pathogenic bacteria in different ways. Firstly, through antimicrobial factors that are released by probiotics. For example, Lactobacillus rhamnosus GG produces compounds that inhibit the growth of several gram-positive and gram-negative bacteria by producing antimicrobial substances such as lactic acid, hydrogen peroxide, and pyroglutamate (Meurman, Antila, Korhonen & Salminen,
Secondly, some probiotics may competitively inhibit adhesion of pathogenic bacteria through resistance to colonization (Goldin, Gorbach, Saxelin, Barakat, Gualtieri & Salminen, 1992). For example, Lactobacillus rhamnosus GG and Lactobacillus plantarum 299V competitively inhibit the attachment of enterohemorrhagic E. coli O157:H7 to HT-29 cells (Mack, Michail, Wei, McDougall & Hollingsworth, 1999). Thirdly, some probiotics may compete for available nutrients thereby limiting the nutrient resources for pathogenic bacteria. For instance, consumption of monosaccharides by a probiotic may reduce the growth of Clostridium difficile, which is dependent on monosaccharides for growth (Wilson & Perini, 1988). Finally, some probiotics have the ability to neutralize toxins produced by pathogenic bacteria. For example, probiotic Saccharomyces boulardii modifies C. difficile toxin A receptor through an enzymatic mechanism thereby neutralizing the toxin effect (Pothoulakis et al., 1993). Figure 22 shows some examples of antibacterial effects of probiotics on pathogenic bacteria.

Figure 22 Probiotic can provide beneficial effects on intestinal cells in different ways. a) Some strains can block pathogen binding to epithelial cells by providing a physical barrier, referred to as colonization resistance. b) Some probiotic strains produce antimicrobial factors. d) Some probiotic compete with pathogen for available nutrients. Adapted and modified from (Gareau et al., 2010).
4.1.2 Influence of probiotics on host intestinal epithelium, mucus production, and barrier function

Probiotics can modulate the activity of gut epithelial cells through different ways. For example, some probiotics can prevent the development of a leaky intestine by enhancing the activation of tight junction proteins (Resta-Lenert & Barrett, 2003). Also, probiotic bacteria such as Lactobacillus rhamnosus GG can prevent inflammation and programmed cell death of the lining intestinal epithelial cells (Yan & Polk, 2002). Furthermore, by enhancing mucous production (mucins MUC2 and MUC3 mRNA expression) in host epithelium, some probiotics can give protection against adhesion of pathogenic bacteria, by increasing and enhancing the expression of intestinal enzymes (Buts, De Keyser & De Raedemaeker, 1994; Jahn et al., 1996) (figure 23).

![Diagram showing the influence of probiotics on host intestinal epithelium](image)

**Figure 23** Some strains can prevent the pathogen from binding to epithelial cells by creating a mucus barrier. This is done by inducing the release of mucus from goblet cells. Adapted and modified from (Gareau et al., 2010).

4.1.3 Influence of probiotics on immune responses

Probiotics can impart their beneficial effect by modulating the immune system, in at least three ways

1. Modulations of humoral immune response - Many studies have demonstrated the capability of many probiotic agents in inducing a specific antibody response. For instance, *L. casei* and *L. plantarum* stimulate rotavirus specific IgA and *E. coli* O157:H7 specific IgA antibodies against rotavirus and *E. coli* infection, respectively (Kaila, Isolauri, Saxelin, Arvilommi & Vesikari, 1995; Conlan & Perry, 1998). Also,
in animal models, ingestion of *Bifidobacterium bifidum* significantly increases the number of immunoglobulin (IgM, IgG, and IgA) secreting cells in mesenteric lymph nodes and spleen (Buts, Bernasconi, Vaerman & Dive, 1990; Park et al., 2002).

2. Modulation of the innate immune response - Probiotic bacteria appear to modulate the non-specific immune response differently in healthy and hypersensitive subjects: in healthy subjects probiotics can be immunostimulatory; whereas, in milk-hypersensitive subjects, probiotics can down-regulate the inflammatory response (Isolauri, Sütas, Kankaanpää, Arvilommi & Salminen, 2001). Probiotics strains, such as *E. coli* Nissle 1917 stimulate the intestinal innate immune system through up-regulation of antimicrobial peptides (e.g., human beta defensin 2) (Wehkamp et al., 2004) and *S. boulardii* activates the complement system (Buts et al., 1994).

Cytokine production: The release of cytokines from cells can be specific for each probiotic strain, i.e., not all probiotic strains have similar immuno-modulating properties. Indeed, different species of *Lactobacilli* exert very different dendritic cell activation patterns and the Th1, Th2, Th3 response of dendritic cells in the intestine can be modulated according to the composition of the gut microflora, including the ingested probiotics (Christensen, Frokiaer & Pestka, 2002). For example, Kwon et al showed that a mixture of five probiotics (*L. acidophilus*, *L. casei*, *L. reuteri*, *B. bifidum*, and *S. thermophilus*) induced an activation of CD4+Foxp3+ regulatory T cells (T_{reg}) in healthy mice. This conversion of T cells into Foxp3+ T_{reg} was directly mediated by regulatory dendritic cells (rDCs) that express high levels cytokines including IL-10 and TGF-β (Kwon et al., 2010). Figure 24 illustrates the effect of probiotics on cytokine production.
3. Modulation of signalling pathways inside the cell: The protective effects of probiotics may be mediated by their ability to modulate the host intracellular signalling pathways. For instance, *S. boulardii* prevents enterohemorrhagic *E. coli* infection by interfering with the transduction pathways that control tight-junction structure as well as by inhibiting NF-κB and MAPK signalling pathways leading to the production of IL-8 (Dahan, Dalmasso, Imbert, Peyron, Rampal & Czerucka, 2003). Similarly, a study by Petrof et al demonstrated that probiotics inhibit the pro-inflammatory NF-κB pathway and trigger the expression of heat shock proteins that protect intestinal cells. The inhibition of NF-κB and the increased expression of heat shock proteins may account for the anti-inflammatory and cytoprotective effects reported for probiotics and may be a novel mechanism of microbial-epithelial interaction (Petrof et al., 2004). Figure 25 illustrates the modulation of signalling pathway by probiotics.
Figure 25 Some probiotics (or their products) may prevent an innate immune response by inhibiting (or activating) NFκB in macrophages and dampening (or priming) the host immune response by influencing the production of IL-8 and subsequent recruitment of neutrophils to sites of intestinal injury. Adapted and modified from (Gareau et al., 2010).

4.2 Transcriptomics in probiotic research

In probiotics research, transcriptomics can facilitate further insight into the complex crosstalk between the probiotics, the commensal microflora and the host gut system (O’Flaherty & Klaenhammer, 2010). Christensen et al showed that *Lactobacilli* differentially modulate the expression of cytokines and the maturation of surface markers in murine dendritic cells suggesting that different LAB strains have the capacity to modulate dendritic cells differentially (Christensen et al., 2002).

The transcriptomics response of host cells to probiotics has revealed that many cell types, and cell signalling pathways are involved in modulating the complex crosstalk between probiotics and host cells. Di caro et al showed that administration of *L. rhamnosus* GG in the mucosa of the healthy male patients for six months induced differential gene expression related to cellular process including immunomodulation, apoptosis, cell-cell signalling, cell cycle and cell adhesion (Di Caro et al., 2005).
Sonnenburg et al. examined the strain-specific effects of *B. thetaiotamicron* and *B. longum* in the innate immune system. The study found that *B. thetaiotamicron* modulated/induced a TNF-α response whereas *B. longum* modulated IFN-γ response. In addition, when both strains were administered not only TNF-α but also the IFN-γ response was induced in the host GIT (Sonnenburg, Chen & Gordon, 2006). In a more recent study, van Baarlen et al. found that administration of *L. plantarum* WCFS1 induced gene expression related to immune tolerance pathways at the mucosal surface of healthy humans. Further analysis of the effect of administrating logarithmic, stationary, and dead cells showed distinct expression patterns: logarithmic phase cells induced a response related to metabolic functions and stationary phase cells induced NF-kB and JUN transcription factors that are both involved in the establishment of immune tolerance. These results support the role and demonstrate the potential of probiotics in modulating the function of GIT (van Baarlen et al., 2009). The transcriptomic approach has also been used to study the effect of bacteriocine production by *L. salivarius* UCC118 and its anti-infective mechanism (Corr, Li, Riedel, O'Toole, Hill & Gahan, 2007).

The analysis of the interaction of pathogenic bacteria or probiotic with host cells at the molecular and cellular levels has become a major research area in recent years. Microarrays provide the advantage to investigate the expression of thousands of genes, and the use of databases and bioinformatics software provides an effective way of managing complex and high-volume data. Bioinformatics also assists in identifying patterns of gene expression changes and the corresponding molecular pathways involved. Functional analyses can quickly reveal key gene networks that direct these functions. These networks can, in turn, be correlated with relevant physiological properties. Blood cell transcriptomics could be an effective tool for diagnostic purpose to evaluate the safety aspects of potential probiotics.

Thus, transcriptomics studies may give further insights into strain specific information and the molecular events taking place as a result of the action of probiotic.

### 4.3 *Lactobacillus gasseri* K7 - a potential probiotic

*Lactobacillus gasseri* is known to be the most dominant species among the *L. acidophilus* group of LAB found in the human GIT (Saito, 2004). The *Lactobacillus gasseri* strain K7 (*L. gasseri* K7) is a human isolate (Matijasic & Rogelj, 2000) evaluated for its potential probiotic
properties in vitro in cell cultures and in vivo in animal models (Matijasic, Mojca, Metoda & Rogelj, 2006; Rogelj & Matijasic, 2006):

1. **In vitro** studies with *L. gasseri* K7 have shown its ability to adhere to Caco-2 cells effectively in a concentration dependent manner (Matijasic, Mojca & Metoda, 2003).

2. *L. gasseri* K7 survives in the GIT, and, at least temporary, colonizes the intestinal mucosa of conventional and gnotobiotic pigs (Matijasic, Stojkovic & Rogelj, 2006).

3. **In vitro** studies using a human macrophage cell line (THP-1) show that *L. gasseri* K7 does not evoke higher secretion of pro-inflammatory cytokines (IL-8, IL-6, TNF-α, IL-1β, IL-12) and even has an anti-inflammatory effect since it lowers the concentration of inflammatory cytokines induced by LPS from *E. coli* (Hacin et al., unpublished data).

4. *L. gasseri* K7 is effective in reducing *E. coli* adhesion to Caco-2 enterocytes thereby showing its competition ability in vitro (Matijasic et al., 2006; Rogelj & Matijasic, 2006). Similarly, in vivo models using conventional and gnotobiotic pigs have also shown that oral administration of K7 (5 × 10¹⁰ cfu per piglet/day) to gnotobiotic pigs infected with enterotoxigenic *E. coli* can diminish the severity of infection (Rogelj & Matijasic, 2006).

5. *L. gasseri* K7 produces bacteriocine, namely gassericin A and gassericin B. These bacteriocins have a wide range of inhibitory activity against different strains of gram positive bacteria (Matijasic & Rogelj, 2000).

6. The safety of *L. gasseri* K7 was evaluated in animal models by clinical observation of inflammatory changes and morphological analysis of the intestine. These studies show no negative effect of *L. gasseri* K7 on health and condition of piglets (Rogelj & Matijasic, 2006).

These studies highlight that *L. gasseri* K7 possesses potential probiotic properties and that it has antimicrobial and immunomodulatory effects. As mentioned before, probiotic strains differ in their immune modulatory activity and extensive research is needed to demonstrate
their safety. Thus, the demonstration of an immunomodulatory effect is one of the necessary steps to identify strains with potential benefits to human health. These preliminary tests are usually performed using cell cultures and animal models. In particular, mice, rats, and pigs are often used as models for studying the safety and effectiveness of probiotics intended for human use. The mouse model was used in the second study to demonstrate the antimicrobial and immunomodulatory effect of L. gasseri K7 in mice challenged with E. coli because the murine species is well understood and characterized genetically and similar to humans in genome size and structure, i.e. it share 99% sequence homology with the human genome (Peters, Robledo, Bult, Churchill, Paigen & Svenson, 2007).

Considerable differences exist in the biological activities, doses and composition between different probiotic preparations (Boyle et al., 2006). For example, Wagner et al studied the effects of 4 different probiotic species (L. reuteri, L. acidophilus, L. rhamnosus GG, and B. animalis) in preventing colonization and sepsis with Candida albicans in mice (Wagner et al., 1997). They found all strains to be protective, but there were significant differences in efficacy and a great diversity of immune effects in terms of antibody and proliferative responses to C. albicans and intestinal inflammatory cell infiltration.

Further studies are necessary to increase our understanding of how probiotic agents produce a beneficiary effect on the host as different strains of probiotic bacteria may work by distinctly different mechanisms. It is also important to recognize that in vitro effects of a probiotic may display opposite behavior in vivo (Christensen et al., 2002).

In the mouse study presented in Chapter 4, our objective was to analyze the immunomodulatory and antimicrobial effect of the L. gasseri K7 in an in vivo model using mice challenged with E. coli O157:H7 infection. Global gene expression analysis was carried out using high-throughput microarray technology.


Hayes, M., Stanton, C., Slattery, H., O'Sullivan, O., Hill, C., Fitzgerald, G. F., & Ross, R. P. (2007). Casein fermentate of Lactobacillus animalis DPC6134 contains a range of


112


Rimbach, G., & Minihane, A. M. (2009). Nutrigenetics and personalised nutrition: how far have we progressed and are we likely to get there? *Proc Nutr Soc.*, 68(2), 162-172.


REFERENCE


Manuscripts
CHAPTER 2 – HUMAN STUDY

Postprandial blood cell transcriptomics in response to the ingestion of dairy products by healthy individuals

Francina M Sagaya\textsuperscript{1,2}, Richard F Hurrell\textsuperscript{2} and Guy Vergère\textsuperscript{1}* 

\textsuperscript{1}Agroscope Liebefeld-Posieux Research Station ALP, Berne, Switzerland
\textsuperscript{2}Human Nutrition Laboratory, Institute of Food, Nutrition and Health, ETH Zurich, Zurich, Switzerland

*Corresponding author

Email addresses:

FMS: francina.sagaya@alp.admin.ch  
RFH: richard.hurrell@ilw.agrl.ethz.ch  
GV: guy.vergeres@alp.admin.ch
Abstract

**Background**: ‘Omics’ technologies now allow a holistic investigation of the interaction of nutrients with the human organism at the molecular level. Genome-wide transcriptomic screening of dairy products in human nutritional intervention studies may thus contribute to the assessment of the properties of these products, in particular of their immunological and metabolic properties. The aim of the study was to investigate the genome-wide postprandial effects of single servings of milk and yogurt on gene expression in the blood cells of human subjects and to identify the downstream physiological processes regulated by the differentially expressed genes. We conducted a randomized, controlled, single blinded, crossover study on 6 healthy male individuals. After an overnight fast, 540 grams of milk or yogurt was ingested by the subjects. Blood samples were collected before (0h) and after (2h, 4h, 6h) ingestion and the blood cell transcriptome was analyzed using a linear kinetic analysis that increases the statistical power of the study.

**Results**: Ingestion of milk and yogurt resulted in the identification of 576 and 625 transcripts comprising 495 and 499 differentially expressed genes (FDR<12.5%), respectively. Pathway analysis indicated that the ingestion of milk and yogurt induces similar postprandial physiological responses. In particular, genes involved in protein biosynthesis and mitochondrial activities, such as oxidative phosphorylation, followed a biphasic kinetics being down-regulated at 2h and up-regulated at 6h whereas the opposite kinetics were observed for apoptotic and inflammatory processes during the same time frame. Also, the human cell transcriptome appeared to be specifically modulated by the ingestion of bovine milk, a property that was further modified when milk was fermented by yogurt.

**Conclusions**: A nutritional intervention study with a crossover design and a kinetic analysis of the postprandial blood cell transcriptome of human healthy subjects having ingested dairy products allows the identification of statistically significant and biologically relevant processes with precision. In particular, a decrease in expression of genes related to inflammation at 6h compared to 0h suggests a potential role for dairy products in the management of chronic inflammatory diseases.
Background

Blood cell transcriptomics is emerging as a powerful approach to capture coordinated and subtle changes induced by nutrients in humans (1-3). This approach can complement the classical nutritional intervention studies. The identification of clinical endpoints or biomarkers that change in a statistically-significant manner often require relatively long term intervention which are often incompatible with the limited resources usually available for exploratory research. As cells in the human organism respond specifically and sensitively to the ingestion of food by inducing gene expression within a few hours of the nutritional stimuli, postprandial gene expression could deliver preliminary information on the early molecular events induced by nutrients. In classical nutrition intervention studies, most of the information obtained so far have been gathered using a hypothesis-driven analytical strategy that focuses on the measurement of a restricted set of pre-selected biomarkers, thus precluding a comprehensive analysis of the underlying physiological processes. On the other hand, genome-wide transcriptomics allow a holistic investigation of the effects of bioactive components, including complex food matrices, on cellular processes. Furthermore, insight into the metabolic networks affected by the nutrients is hampered due to the restricted access of researchers to cellular material, in particular tissue biopsies, in human subjects. In blood cell transcriptomics, white blood cells can easily be recovered from human subjects, respond metabolically to nutritional stimuli and are important mediators of the immune system what makes them attractive cellular targets for preventive nutritional research, particularly in the context of chronic inflammatory diseases. Indeed, blood cell transcriptome has been recently used to deliver information on the molecular mechanisms underlying dietary-induced physiological changes (1-4), in particular immunological processes (1).

Dairy products such as milk and yogurt are important components of Western diets. These products not only provide essential macronutrients, vitamins, trace elements and minerals but also contain bioactive components that may act synergistically to modulate physiological processes important to the maintenance of health (5). In this context, the impact of dairy products has been discussed, and debated, in relation to the development or the prevention of cancer (6), allergy (7), hypertension (8), infection (9), and obesity (10). Significant attention has also been paid to the immunomodulatory properties of dairy products (11-14). With respect to these properties, we have conducted a nutritional intervention study in healthy
subjects to investigate how milk and yogurt influence the postprandial transcriptome of blood cells.

Results and Discussion

Identification of differentially expressed genes after the ingestion of milk and yogurt

This transcriptome study assesses kinetic changes in gene expression after the ingestion of dairy products in humans. A repeated measures ANOVA, considering the gene expression analyses at the time points 2h, 4h and 6h relative to 0h as repeated measures, identified 70 genes after the ingestion of milk (M group) and 5 genes after the ingestion of yogurt (Y group) that are differentially expressed using a 12.5% FDR threshold. A quadratic contrast, assuming a quadratic function for the changes in gene expression between time points 2h, 4h, and 6h, did not lead to the identification of differentially expressed genes. However, a linear contrast, assuming a linear change of gene expression between 2h and 6h, identified 495 differentially expressed genes (576 transcripts, p ≤ 0.002) in the M group and 499 differentially expressed genes (625 transcripts, p ≤ 0.002) in the Y group. Both groups shared 11% of their differentially expressed genes. Out of 70 significant genes selected from the repeated measures ANOVA, 65 genes were also found significant in the linear contrast test. The list of all differentially expressed genes is available in Additional file 1.

Kinetic analysis of the genes in the M and Y groups

Figure 1 illustrates the average kinetics of differential gene expression in the form of boxplots showing the average SNOMAD Z-scores as a function of time for the genes in the M and Y groups that had either a positive or a negative linear contrast value. For all four groups of genes, the kinetics showed a linear trend between the time points 2h and 6h in agreement with the identification of these genes by significant linear contrasts. Despite this linear trend between 2h and 6h, the kinetics of gene expression is not linear throughout the entire postprandial phase observed in our study, i.e. between 0h and 6h. Indeed, one-sample Wilcoxon signed rank tests on the median Z-scores of the four groups of genes at the time points 2h, and 6h confirmed the visual observation that the median values at 2h (M2/M0, Y2/Y0) were either negative (for the group of genes with positive linear contrasts) or positive (for the group of genes with negative linear contrasts) and significantly different from zero (p value < 0.05) whereas the median values at 6h (M6/M0, Y6/Y0) followed the inverse pattern and were either positive (for the group of genes with positive linear contrasts) or negative (for
the group of genes with negative linear contrasts) and significantly different from zero (p value < 0.05).
(A) Genes with downUP kinetics

![Boxplot of the downUP genes in the M and Y groups](image)

(B) Genes with upDOWN kinetics

![Boxplot of the upDOWN genes in the M and Y groups](image)

Figure 1 - Boxplot of the downUP and upDOWN genes in the M and Y groups. Analysis of the SNOMAD Z-score ratio for the time point comparisons of all genes in the M (M2/M0, M4/M0, M6/M0) and Y (Y2/Y0, Y4/Y0, Y6/Y0) groups. Y0/M0 represents the Z-score ratio of two 0h times from the cross over design which is used as a reference in this analysis. The relative SNOMAD Z-score ratio is shown on the y axis. The boxplots show the maximum value, lower quartile, median, upper quartile, and minimum values observed for each time point in the groups indicated. (A) Boxplot of downUP genes in the M (307 transcripts) and Y (394 transcripts) groups. (B) Boxplot of upDOWN genes in the M (269 transcripts) and Y (231 transcripts) groups.

Taken together the statistical analysis of the data at the various data points showed that each of the four sets of genes followed biphasic kinetics. For the set of genes with a positive linear contrast in the M and Y groups we observed an initial transient down-regulation between 0h and 2h followed by a more pronounced and sustained up-regulation between 2h and 6h, the 6h values being higher than the 0h values (Figure 1A). We refer to this biphasic kinetics in
the text as ‘downUP kinetics’. The opposite effect was observed for the set of genes with a negative linear contrast in the M and Y groups as these genes first showed a transient increase in expression between 0h and 2h followed by a more pronounced and sustained decrease between 2h and 6h, the 6h values being lower than the 0h values (Figure 1B). We refer to this biphasic kinetics later in the text as ‘upDOWN kinetics’.

Several, non-exclusive, explanations can be put forward to account for the biphasic kinetic pattern: (i) Homeostatic response of the organism that responds to the initial diet-induced stress by restoring the original equilibrium of the biological system. Of note, in apparent disfavour to this hypothesis, the second phase of the kinetics proceeds beyond the equilibrium point at time 0h; (ii) Superimposition of different kinetics of gene expression arising from different types of blood cells; (iii) Synergistic and antagonistic properties of different nutrients in dairy products; (iv) Kinetic effects resulting from the digestion and absorption processes of the dairy products in the gastrointestinal tract. Further experiments need to be conducted to differentiate between these hypotheses. Meanwhile, our findings highlight the importance of conducting kinetic studies while looking at the expression of genes resulting from the ingestion of nutrients as the measurement of single time points only identifies a limited number of genes and leads to an incomplete assessment of the biological response.

**Correlation analysis of the linear contrasts for the sets of differentially expressed genes in the M and Y groups**

A comparison of the linear contrast values of the significant genes in the M group to the linear contrast values of the same genes after the ingestion of yogurt showed a statistically significant positive correlation of 0.90 between the two data sets (Figure 2A).
Figure 2 - Spearman rank correlation analysis using linear contrast values. (A): Linear contrast values of all genes of the M group plotted against the corresponding linear contrast values for the same genes after ingestion of yogurt (green open circles). (B) Linear contrast values of all genes of the Y group plotted against the corresponding linear contrast values for the same genes after ingestion of milk (blue crosses). (C) Combined panels (A) and (B).

*Spearman rank correlation coefficient.
Similarly, the linear contrast values of the significant genes in the Y group, when compared to the linear contrast values of the same genes after the ingestion of milk, also showed a statistically significant positive correlation of 0.80 between the two data sets (Figure 2B). Finally, the correlation between the linear contrast values of all the genes differentially expressed in either the M or the Y group reached a value of 0.79 (Fig. 2C).

Milk and yogurt share more similarities in their chemical composition than they differ which is reflected in the observation that the two sets of linear contrasts for the differentially expressed genes were positively correlated. On the other hand, the lower correlation obtained with the differentially expressed genes of the Y group, compared to the correlation obtained with the differentially expressed genes of the M group, is intriguing. As the microbiological transformation of milk to yogurt increases the complexity of its chemical composition, we hypothesize that this difference in the postprandial blood cell transcriptome may reflect changes in the nutritional properties of this product.

**Pathway enrichment profiles of genes in the M and Y groups**

Figure 3 shows a process network analysis of the genes differentially expressed in response to the ingestion of milk and yogurt. Overall most of the biological processes were identified within the group of genes having a negative linear contrast. In particular, processes mediating the immune response, signal transduction, apoptosis and the inflammatory response contained genes with negative linear contrasts. Analysis of the genes having a positive linear contrast identified only two major processes, namely transcription and translation.

The shape of the functional profiles of the M and Y groups of genes in Figure 3 were globally similar suggesting that milk and yogurt induce a common postprandial response in humans. Interestingly, despite the fact that the M and Y groups contained a similar number of differentially expressed genes, a larger number of processes reached statistical significance in the M group than in the Y group. This finding suggests a quantitative, rather than a qualitative, difference in the postprandial response of the subject to the ingestion of milk and yogurt.

We also conducted a pathway enrichment analysis of the set of genes with statistically significant linear contrasts in the M and Y groups using GenMAPP (Table 1).
Table 1 - Overrepresented functional groups in the total set of differentially expressed genes in the M and Y groups

Overrepresentation analysis was performed in GenMAPP 2.0 by comparing differentially expressed genes in the M and Y groups to all genes on the microarray. GO, gene ontology functional groups; CC, cellular component; BP, biological process; MF, molecular function. A, number of genes measured; B, number of genes in GO group; C, number of genes changed; D, Z score with p ≤ 0.05 (permutated P).

<table>
<thead>
<tr>
<th>GO</th>
<th>Functional groups containing genes with positive linear contrasts</th>
<th>Milk</th>
<th>Yogurt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>BP</td>
<td>Biosynthetic process</td>
<td>1261</td>
<td>1358</td>
</tr>
<tr>
<td>BP</td>
<td>RNA processing</td>
<td>403</td>
<td>424</td>
</tr>
<tr>
<td>BP</td>
<td>Lymphocyte differentiation</td>
<td>86</td>
<td>93</td>
</tr>
<tr>
<td>MF</td>
<td>RNA Binding</td>
<td>636</td>
<td>673</td>
</tr>
<tr>
<td>CC</td>
<td>Ribosome</td>
<td>217</td>
<td>249</td>
</tr>
<tr>
<td>CC</td>
<td>Mitochondrial part</td>
<td>359</td>
<td>342</td>
</tr>
<tr>
<td>CC</td>
<td>Spliceosome</td>
<td>128</td>
<td>132</td>
</tr>
<tr>
<td>CC</td>
<td>Ribonucleoprotein complex</td>
<td>446</td>
<td>485</td>
</tr>
<tr>
<td>CC</td>
<td>Structural constituents of ribosome</td>
<td>195</td>
<td>226</td>
</tr>
<tr>
<td>CC</td>
<td>Small ribosomal subunit</td>
<td>59</td>
<td>64</td>
</tr>
<tr>
<td>CC</td>
<td>Large ribosomal subunit</td>
<td>65</td>
<td>68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GO</th>
<th>Functional groups containing genes with negative linear contrast</th>
<th>Milk</th>
<th>Yogurt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>BP</td>
<td>Signal transduction</td>
<td>3006</td>
<td>3464</td>
</tr>
<tr>
<td>BP</td>
<td>Apoptosis</td>
<td>715</td>
<td>716</td>
</tr>
<tr>
<td>BP</td>
<td>Inflammatory response</td>
<td>284</td>
<td>303</td>
</tr>
<tr>
<td>BP</td>
<td>Immune response</td>
<td>590</td>
<td>714</td>
</tr>
<tr>
<td>BP</td>
<td>Cell motility</td>
<td>420</td>
<td>416</td>
</tr>
<tr>
<td>BP</td>
<td>Acute inflammatory response</td>
<td>67</td>
<td>75</td>
</tr>
<tr>
<td>BP</td>
<td>Cytokine production</td>
<td>111</td>
<td>110</td>
</tr>
<tr>
<td>BP</td>
<td>Cell surface receptor linked signal transduction</td>
<td>1393</td>
<td>1745</td>
</tr>
<tr>
<td>BP</td>
<td>Cell adhesion</td>
<td>657</td>
<td>688</td>
</tr>
<tr>
<td>BP</td>
<td>Angiogenesis</td>
<td>139</td>
<td>131</td>
</tr>
<tr>
<td>CC</td>
<td>Cytoplasmic vesicle</td>
<td>256</td>
<td>256</td>
</tr>
</tbody>
</table>

At the level of the biological processes this analysis provided us with similar results than obtained with the process network analysis shown in Figure 3. In addition, the GenMAPP analysis with the genes having a positive linear contrast revealed an enrichment of cellular components associated with ribosomal and mitochondrial activity while the processes containing genes with negative linear contrasts were localized to the extracellular matrix and to cytoplasmic membrane-bound vesicles. The data thus indicate that the protein biosynthesis machinery and the mitochondrial functions are up-regulated at 6h whereas immunomodulatory functions involving signal transduction are down-regulated at 6h.
Figure 3 - Overview of biological processes in blood cell gene expression profiles induced by milk and yogurt ingestion. Barplot showing the processes with negative and positive linear contrasts for the M and Y groups resulting from a network analysis performed using MetaCore version 5.4. A -log(p value) >2 is considered as statistically significant (dotted lines).
Figure 4 - Genes statistically changed (FDR < 12.5%) after ingestion of milk and yogurt in various overrepresented functional processes. Heatmap of genes associated with multiple groups are indicated with an asterisk (*) appended to the gene name. Compared to 0h, the green and red colours represent down-regulated and up-regulated genes, respectively. The coloured-graded bar gives the z score ratio. The distribution of the genes into the functional processes was obtained by GenMAPP pathway analysis.
Figure 5 graphically presents the linear contrasts of the genes identified in Table 1 that are differentially expressed in either the M or the Y group and that belong to the functional groups ‘mitochondrial part’, ribosomes’, and inflammatory response’. This graph again highlights the similarity in the postprandial response of the subjects to milk and yogurt. More strikingly, the linear contrasts of the genes belonging to the same functional group almost exclusively segregate in the same quadrants of Figure 5, i.e. positive values for ribosomal and mitochondrial genes, and negative values for inflammatory genes.

**Figure 5 - Linear contrasts in the M and Y groups for genes belonging to selected functional groups.** All genes of the M (grey circles) and Y (grey crosses) groups; All mitochondrial genes in M and Y groups are shown as pink rhombus (24 out of 25 genes and 13 out of 14 genes have positive linear contrasts in the M and Y groups, respectively); All ribosomal genes are shown as cyan triangles (27 out of 28, genes and 34 out of 35 genes have positive linear contrasts in the M and Y groups, respectively); All inflammatory genes are shown as red squares (12 out of 13 genes and 11 out of 13 genes have a negative linear contrast in the M and Y groups, respectively).

**Kinetic and functional analysis of ribosomal genes**

The left panels of Figure 6A (M group) and 6B (Y group) illustrate the downUP kinetics of genes involved in protein biosynthesis, a group of genes almost exclusively composed of genes coding for ribosomal proteins. The largest difference in relative gene expression was observed between time points 2h and 6h. An analysis of the data by hierarchical clustering of the time point comparisons also showed, for both sets of ribosomal genes in the M (right panel of Figure 6A) and Y (right panel of Figure 6B) groups, that the 2h and 6h time points maximized the differences in relative gene expression.
Down-regulation of ribosomal genes has been observed in healthy humans two hours after the consumption of a protein-rich breakfast (2). This down-regulation has been interpreted as a response of the organism to protein overload. In our study we extend these observations as we observed biphasic downUP kinetics for the ribosomal genes. The pronounced up-regulation following the transient decrease at 2h could be attributed to the presence of proteins in the dairy products, whose amino acids serve as building blocks for the synthesis of endogenous proteins (15). In particular, whey proteins are rich sources of branched chain amino acids, such as leucine, and play a pivotal role in the initiation of protein synthesis in muscle (16-19). One can thus speculate that the organism responds to the protein overload by first inhibiting protein synthesis, but that dairy products possess intrinsic nutritional properties that eventually lead to the activation of protein synthesis. Indeed, dairy proteins are more efficient in sustaining protein synthesis in comparison to soy proteins in human muscle cells (20). This activity may be of importance in the management of muscle mass in physically active persons and the elderly. However, this remains to be demonstrated in future studies (21, 22).
Figure 6 - Boxplot and hierarchical clustering of the ribosomal genes in the M and Y groups. Analysis of the SNOMAD Z-score ratio for the time point comparisons of the 26 ribosomal genes in the M group (M2/M0, M4/M0, M6/M0) and of the 31 ribosomal genes in the Y group (Y2/Y0, Y4/Y0, Y6/Y0). Y0/M0 represents the Z-score ratio of two 0h times from the cross over design. Left panel: boxplot showing the distribution of the Z-score ratio for each time point comparison. Right panel: hierarchical clustering of the time point comparisons using Euclidean distance measure. The red colour indicates up-regulation of gene expression relative to 0h. The green colour indicates down-regulation of gene expression relative to 0h. The dendrogram illustrates that 2h and 6h are the two farthest clusters.
Kinetic and functional analysis of mitochondrial genes

The group of proteins with a mitochondrial function (e.g. genes involved in oxidative phosphorylation: \textit{NDUFB7, NDUFB9, NDUFS3,} and \textit{NDUFS4} which are subunits of complex I, \textit{ATP synthase; UQCR; cytochrome c}) also follows downUP kinetics (not shown).

The similarity in downUP kinetic patterns observed with ribosomal and mitochondrial genes is in line with the role of oxidative phosphorylation as a provider of metabolic energy for the protein biosynthetic process (23). As carnitine and calcium stimulate oxidative phosphorylation (24, 25), the presence of these nutrients in dairy products makes them good candidates for the induction of mitochondrial gene expression observed in our study. Finally, on another line of interest, the enzymatic activity of complex I in blood cells responds within a few days to the re-feeding of malnourished patients in a more sensitive manner than validated nutritional markers such as albumin (26). Our observation that changes in the expression of 33 mitochondrial genes, including complex 1, can be measured within a few hours of feeding substantiates the use of mitochondrial nutritional markers in clinical nutrition.

Kinetic and functional analysis of inflammatory genes

The left panels of Figure 7A (M group) and 7B (Y group) show the upDOWN kinetics of genes involved in the inflammatory response in the form of boxplots (average kinetics for the group of genes) and illustrates that the largest difference in relative gene expression was observed between the time points 2h and 6h. An analysis of the data by hierarchical clustering of the time point comparisons also showed, for both sets of inflammatory genes in the M (right panel of Figure 7A) and Y (right panel of Figure 7B) groups, that the 2h and 6h time points maximized the differences in relative gene expression. Of note, the upDOWN kinetics of the inflammatory group of genes appeared to be more pronounced in response to the ingestion of milk than of yogurt.
Figure 7 - Boxplot and hierarchical clustering of the inflammatory genes in the M and Y groups. Analysis of the SNOMAD Z-score ratio for the time point comparisons of the 12 inflammatory genes in the M group (M2/M0, M4/M0, M6/M0) and of the 11 ribosomal genes in the Y group (Y2/Y0, Y4/Y0, Y6/Y0). Y0/M0 represents the Z-score ratio of two 0h times from the cross over design. Left panel: boxplot showing the distribution of the Z-score ratio for each time point comparison. Right panel: hierarchical clustering of the time point comparisons using Euclidean distance measure. The red colour indicates up-regulation of gene expression relative to 0h. The green colour indicates down-regulation of gene expression relative to 0h. The dendrogram illustrates that 2h and 6h are the two farthest clusters.
Key players in the induction of inflammation, such as Rel-A (p65 component of NF-κB heterodimer), IL6, IL1β, TLR2, TLR4, and IL8RB were differentially regulated after the ingestion of dairy products. Figure 8 shows a GeneGO Metacore map illustrating the Toll like receptor (TLR) signalling cascade and all the genes in these pathways which were differentially expressed with a p value <0.05 in either the M or Y group. With the exception of TRAF6, all other genes had a negative value in the linear contrast indicating down-regulation at 6h. These genes code for proteins that cover most of the signalling cascade from cell surface receptors (CD14; TLR2, TLR4, TLR6), intracellular mediators (IRAK1, IRAK1/2, MEK3, NIK, TLR8, TOLLIP, TRAF6), nuclear transcription factors (NF-κB), and finally extracellular effectors (IL6, IL8).

TLRs are major mediators of the inflammatory signalling pathways (27). Although TLRs are primarily involved in the recognition of pathogenic motifs (e.g. recognition of lipopolysaccharide (LPS) from gram-negative bacteria by TLR4) studies suggests that TLR4 is activated by saturated fat and antagonized by unsaturated fat thereby linking nutrition to inflammation and chronic diseases (28). Together with other components of the TLR signalling pathway, the activity of TLR2 and TLR4 are modulated by the addition of milk to epithelial cells in culture (29). More relevantly, a nutritional intervention study in human healthy subjects has demonstrated postprandial activation of TLR2 and TLR4 genes and proteins in mononuclear cells in response to a diet enriched in fat and carbohydrates (30). Furthermore, addition of orange juice to the diet counteracted these effects. In that context, our finding that the ingestion of dairy products is accompanied by an overall down-regulation of inflammatory genes at 6h strongly suggests that the biological activity of nutrients in dairy products with a potential for anti-inflammatory activity (e.g. unsaturated fat (1) or whey proteins) prevails over the activity of nutrients with pro-inflammatory activity (e.g. saturated fat) (31, 32). Of interest, several dairy components such as CLA (33), TGF-β, and lactoferrin (34, 35) not only possess anti-inflammatory properties but their activities are enhanced in milk food matrix.
Figure 8 - Genes in Toll like receptor (TLR)-signalling pathway that are differentially expressed after the ingestion of milk and yogurt. All genes indicated with the thermometer symbol represent genes which are significantly expressed with linear contrasts with p-values < 0.05 in milk and yogurt dataset. Black arrows on the gene name indicate that the gene is differentially expressed in the M and Y groups (FDR < 12.5%). The thermometers labelled ‘1’ indicate genes in milk dataset. The thermometers labelled ‘2’ indicate the yogurt dataset. A thermometer with the designated number on the top and blue colour indicates genes with a negative linear contrast; a thermometer with the designated number on the bottom and red colour indicates genes with a positive linear contrast. The magnitude of expression is indicated by the height of the coloured column in each thermometer. Abbreviations for the differentially expressed genes: CD14 (CD14 antigen); IL6 (interleukin 6); IL8 (interleukin 8); IRAK1 (interleukin-1 receptor-associated kinase 1); IRAK1/2 (interleukin-1 receptor-associated kinase 1); MEK3 (mitogen-activated protein kinase kinase 3); NF-kb (nuclear factor of kappa light polypeptide gene enhancer in B-cells); NIK (NF-kappa beta-inducing kinase); TLRs (Toll-like receptors); TOLLIP (Toll-interacting protein); TRAF6 (TNF receptor-associated factor 6).
Kinetic and functional analysis of apoptotic genes

The group of genes involved in the apoptotic process (e.g. *BCL10*, *TP53BP2*, *PDCD6IP*, *BAG5*, *PTEN*) also follows average upDOWN kinetics (not shown). The initial increase in the upDOWN kinetics of genes involved in apoptosis may be due to the production of reactive oxygen species (ROS) resulting from the feeding process (36-38). Indeed, ingestion of lipids and proteins as well as of a mixed meal after an overnight fast induces the generation of reactive oxygen species (ROS) (36, 37). Based on these findings we hypothesize that the transient increase of apoptotic genes at 2h can be attributed to the generation of ROS.

Interestingly, in obese individuals and in ap2-agouti mice, dietary calcium and dairy products suppress ROS generation (13). This observation has been confirmed in a recent human study which showed that dairy components suppress ROS (14). In line with these findings we observed an increased expression at 6h of the genes coding for superoxide dismutase 1 (M: p \( \leq 0.03 \); Y: p \( \leq 0.0003 \)) and thioredoxin 2 (M: p \( \leq 0.002 \); Y: p \( \leq 0.02 \)) and hypothesize that antioxidants, in combination with other components in dairy products (17, 39, 40), promote the subsequent, more pronounced, down-regulation of pro-apoptotic genes at 6h.

Apoptosis is a highly regulated process that plays an important role in the maintenance of tissue homeostasis (41, 42). Apoptotic and inflammatory processes are coordinately activated in response to cellular stress (43-46). For example, NF-kB participates in the activation of inflammation and apoptosis in the presence of stress factors, including ROS. Also superoxide dismutase is involved in the suppression of both inflammation and apoptosis (47). The similar upDOWN kinetics observed with the inflammatory and apoptotic genes may thus reflect this synergistic mode of action.

**Gene set enrichment analysis (GSEA) to identify pathways in the Y group that are differentially regulated in comparison to the M group.**

GSEA, in addition to Genmapp and GeneGO pathway analysis, further supported our above findings that most of the biological processes regulated in the Y group are shared by the M group (Figure 9). However, despite these similarities, GSEA revealed eight pathways in the Y group which were significantly different from the M group. These pathways were namely, ACE2 (Angiotensin converting enzyme 2), AHS (alpha-hemoglobin stabilizing protein), ETS
(macrophage differentiation), INTRINSIC (intrinsic prothrombin activation pathway), MTA3 (metastasis associated 1 family, member 3 pathway), P53 (p53 signalling), VITCB (Vitamin C in the brain pathway) and WNT (wnt signalling). Lack of data with respect to the influence of nutrients in yogurt on these pathways restricts further discussion. However, enrichment of the ACE2 pathway in the Y group provides an interesting observation. The renin-angiotensin system (RAS) is an important pathway for regulating blood pressure in humans. The key players in RAS are ACE1 and ACE2, two enzymes with opposing activities. ACE1 catalyses the conversion of angiotensin I (Ang I) to angiotensin II (Ang II) which is a vasoconstrictor whereas ACE2 catalyses the conversion of Ang II to Ang (1-7), a peptides with vasodilator properties (48, 49). The fermentation process using lactic acid bacteria in milk releases ACE1 inhibitory peptides from casein (50, 51). These peptides inhibit the activity of ACE1 enzymes and may have implication in regulation of blood pressure in humans.

![Venn diagram showing common and differently regulated pathways in the Y group in comparison to the M group. Out of 63 pathways in the Y group, 55 pathways were shared with the M group and only 8 were unique.](image)

**Statistical validity of study**

Despite the modest sample size (the transcriptome of six subjects was finally measured) and the small changes in expression (the fold changes can be deduced from Additional file 1 and Figure 1) we have identified several hundred genes that are differentially expressed in postprandial response to the consumption of dairy products. We attribute this positive outcome to our kinetic strategy in the analysis of the data. Indeed, the power of the F-test on linear trends is higher than the power of the overall repeated measures ANOVA F-test because of the bigger non centrality parameters of the noncentral F-distribution used in power calculation. A more direct consideration with the same conclusion is the comparison of the
error variance used to calculate the F-test statistics for the repeated measures ANOVA with
the error variance used for the F-test statistics of the linear contrasts: in our case, the latter is
only one third of the former. Thus, this reduction of variance clearly means more power for
the F-tests of the linear contrasts compared to the repeated measures ANOVA F-tests.

In light of the low percentage of differentially expressed genes shared between the M and Y
groups (11%), the significant correlation between the linear contrasts of both data sets
suggests that the Benjamini-Hochberg calculation overestimates the true FDR (therefore a cut
off value of 12.5% was selected). Despite this limitation, the sets of differentially expressed
genes used for the functional analyses in the M and Y groups globally lead to the
identification of similar physiological processes and pathways for both dairy products. This
finding therefore validates the selection criteria used for the statistical cut off values.

Intra- and inter-individual variations between the two starting points, though statistically not
important for the study of linear contrasts, were assessed by the Z-scores of the Y0/M0
comparison. Intra-individual variation was considered significant for absolute Z-scores ≥
1.96. Proportions of about 17 % of the whole 43376 transcripts and about 16 % of the 650
differentially expressed transcripts exceeded this limit. The proportions of significant inter-
individual variation were estimated by comparing the means of the six Z-scores to the 0.025-
and 0.975-quantiles of the expected distribution. About 18 % and 14 % for the whole and
differentially expressed transcripts, respectively, were outside these significance limits; a
finding similar to other studies (5, 7). The similar proportions of significant intra- and inter-
individual variation for differentially expressed transcripts and the whole transcripts indicate
the increased sensitivity of the kinetic analysis by linear contrasts compared to common
pairwise t-tests.

A definite asset of the study is the crossover design in which each subject served as his own
control. A combined effect of the kinetic and crossover design is the fact that the error term in
the denominator of the F-Tests is the mean intra-individual variance which does not depend
on the inter-individual variability and therefore allowed us to identify the genes changing
their expression after the ingestion of the dairy products with accuracy. The boxplots in
Figure 1 illustrate this conclusion for all four panels as the mean values of the SNOMAD Z-
score ratios for the comparisons M6/M0 and Y6/Y0 were all statistically different from zero
whereas the Y0/M0 were not. In other words, despite the known inter-individual variability in
human transcriptomic analyses (2), we can identify genes that change their expression within
a few hours of the ingestion of dairy products and that do not vary over a time frame of four days when measured under fasting conditions. Interestingly, as the mean values of the SNOMAD Z-score ratios for the two Y0/M0 comparisons were not significantly different from zero, we also conclude that a carryover effect is not observed and that our study was appropriately designed in that regard.

The postprandial response is characterized by modest changes in the intensity of the expressed genes in human blood cells. Consequently, we observed that the housekeeping genes usually selected for conducting confirmatory RT-PCR analytics vary in their intensity values on the microarrays in a range that is close to the fold change identified for the sets of differentially expressed genes (data not shown). As the normalization procedures used with microarrays are more comprehensive and robust than RT-PCR (52, 53), and as the analysis of our data clearly points to biologically relevant phenomena (see Figure 5), we did not use RT-PCR as a control method to confirm the validity of the microarray data. To take this limitation into account, we conduct most of the discussion of the microarray data at the level of biological processes and pathways being careful not to over-interpret results from the finding of single genes. In that context, a functional analysis of the blood cell transcriptomes induced in response to milk and yogurt should not be attempted at the gene-by-gene level.

**Biological validity of study**

Taking a heterogeneous population of blood cells for a transcriptomic analysis of the data restricts the interpretation of the data as it is not possible to differentially evaluate the specific contribution of each type of blood cells to the postprandial response. On the other hand, this disadvantage can be turned into an advantage as a transcriptomic analysis of whole blood may fulfil the needs of nutrition research by only allowing the detection of universal metabolic phenomena that will eventually be physiology relevant, ignoring more specific effects.

The representativity of our study may be questioned due to the small number of subjects. However, this restriction is induced by the selection criteria (gender, age, BMI) and much less by the number of individuals. In fact, since we have identified many genes that respond to the ingestion of dairy products in a statistically-significant and biologically meaningful manner with a small sample size, this effect is most probably important. The striking functional clustering of the differentially expressed genes in the negative (inflammatory
genes) or positive (ribosomal genes; mitochondrial genes) quadrants of the linear contrasts shown in Figure 5 illustrates this conclusion in a dramatic manner. This finding can only be explained by a biological interpretation of the data that is likely to be universal and not restricted to the six individuals tested.

The biological pathways identified in this study could either be attributed to the direct effects of the dairy products on the blood cell transcriptome of the organism or to endogenous effects that are independent of the ingested dairy products. The gene expression program induced during the fasting phase (e.g. the circadian rhythm or fasting metabolism) may overlap with the feeding phase of the study (54). However, we found that only ~7% of the blood cell genes that are differentially expressed after 6h in response to the dairy products also change their expression during a 48h fasting period (4). We thus conclude that the contribution of the fasting and circadian metabolic processes to our sets of differentially expressed genes is limited. The acid or volume load resulting from the ingestion of 540g of dairy products may also trigger gene expression. To check this possibility we performed a text based search using PubMatrix (55) on the literature database (NCBI PubMed). We uploaded the M and Y group of genes and searched against the following keywords: acid load, acid overload, acidosis, lactic acid, acid challenge, acid base equilibrium, acid base imbalance, gluconic acid. The results from this search (data not shown) did not provide evidence that an acid overload was responsible for the gene expression profile observed in our study. Indeed, the functional analyses of our data strongly indicates that the observed postprandial changes in gene expression directly result from the ingestion of nutrients present in dairy products.

The results of the GSEA analysis on the expression of human genes homolog to bovine milk proteins strikingly illustrate the above conclusion. Indeed, the extent to which proteins or bioactive peptides (after milk processing, digestion and absorption into the circulation) in dairy products, milk in particular, may express their specific biological activity in humans is an intriguing question. Whereas such effects are clearly important in breastfed babies (56), their relevance in human adults is clearly less evident and the subject of more debate. In order to specifically address this point we investigated whether the postprandial genes differentially expressed in the blood cells of our adult subjects were indicative of specific biological properties that may be attributed to the ingested bovine milk proteins. Using the GSEA algorithm (57) we therefore evaluated the distribution of human genes - that are homolog to proteins present in bovine milk – in the list of genes ranked by the FDR-values associated
with their postprandial linear contrasts. We observed that the 112 gene products identified in bovine milk (Additional file 2) clustered towards the top of the milk gene list (e.g. the most differentially expressed genes) with a corresponding p-value of 0.0024. In other words, these 112 genes tended to be more differentially expressed than other genes in the 4x44k array. The statistical relevance of this analysis was further increased when the GSEA analysis was performed with the list sorted first according to the direction of differential expression and then according to the FDR-values associated with their postprandial linear contrasts (up-regulated genes, with positive linear contrast, at the top, and down-regulated genes at the bottom). The GSEA indicated that milk genes tended to cluster towards the end of the list (down-regulated genes, p=0.0094) rather than towards the top (p=0.9925).

In addition to the GSEA analysis, we have identified eight human genes coding for homolog gene products present in bovine milk that have a statistically significant linear contrast in human blood cells after the ingestion of the dairy products (A1BG, ALDH2, HPSE, SERPING1, TLR4, and YWHAB in the M group; HPSE, TLR2, and SERPINA1 in the Y group). Strikingly, genes belonging to the same protein families were separately identified in the M and Y groups (Toll-like receptors: TLR2, TLR4; Serpins: SERPING1, SERPINA1; Heparanase: HPSE), a finding that is highly unlikely to occur by chance. Also, seven genes, out of eight, showed a negative linear contrast (only A1BG, whose function is unknown, had a positive linear contrast). For most of these genes, sufficient data is not available in the literature to support the development of hypotheses on the potential biological activity of the ingested bovine proteins by humans. However, the presence of membrane-bound and soluble forms of TLR2 and TLR4 is well documented in bovine and human milk (29, 58, 59). Also, soluble TLR2 in human breast milk modulates the cellular response to bacterial components by regulating the activity of the TLR co-receptor CD14 (59).

The most straightforward interpretation for the above results is a negative feedback postprandial regulation of human blood cell genes by bovine milk proteins or peptides. Keeping in mind the sequence homology between the human and bovine genomes, our findings lead us to propose that the bioactivity of dairy products crosses the barriers of species (Homo sapiens - Bos taurus) and life cycles (suckling offspring – adult consumer) to deliver nutritional properties that extend beyond their macronutrient composition.
Conclusions

Our study demonstrates that a blood cell transcriptomic approach to investigate the postprandial properties of foods and nutrients in humans is technically feasible, delivers meaningful biological data, and has a high potential as a discovery tool for nutrition research. More specifically, our results show that the ingestion of milk and yogurt globally induces a similar postprandial response in healthy individuals as a result that would be expected from two products based on the same raw material. In particular, genes involved in protein biosynthesis and energy metabolism are stimulated at 6h whereas inflammatory and apoptotic pathways are down-regulated at 6h. These findings may have implications for the role of dairy products in the dietary management of aging and chronic diseases (60). To give more support to this conclusion, the robustness of our data should be complemented by studies in which a selection of relevant biomarkers (e.g. proteins derived from our list of differentially expressed genes) are measured in healthy subjects and/or in subjects with an underlying imbalance of homeostasis, e.g. with subclinical chronic inflammation. Indeed, a recent study in obese and overweight individuals, already provides support for the anti-inflammatory properties of dairy products (14).

Despite overall similarities between the M and Y groups, the Y group showed a few specific differences in comparison to the M group. One of the striking differences was a significant down-regulation of the ACE2 pathway in the Y group, a pathway that regulates blood pressure in human and that is already discussed in the literature with respect to antihypertensive peptides present in fermented dairy products.

Our data point to the concept of postprandial stress, a normal response of the organism to the ingestion of food that expresses itself in the activation of metabolic, inflammatory and oxidative pathways (61). The pathways identified in our study are in line with this concept. In that context, our kinetic analysis of the transcriptomic data provides us with a quantitative tool with a potential for a better evaluation of the nutritional and metabolic properties of food products and to establish new hypothesis on these properties. Once established, postprandial blood cell transcriptomics in human will eventually be used to monitor the consequences of transforming food, in particular dairy products, on human metabolism and health.
Methods

Study design and subjects

This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the ethical committee of the ETH Zurich, Switzerland.

Interested healthy male volunteers from the student population filled a questionnaire on lifestyle, medical history and dietary habits and reported a western lifestyle. Smokers and subjects who were on a medically-prescribed diet were excluded from participation. Also, subjects who were on medication or who reported metabolic, gastrointestinal disorders, or a history of medical or surgical events that could affect the study outcome were excluded, as well as subjects who had donated blood less than 4 months before the beginning of the study. Eight healthy young males were recruited and written consent was obtained from all subjects. The numbers of subjects were kept small as this was an initial exploratory study. Two subjects had to be excluded during the study because of the abnormal neutrophil count and low RNA yield (see below). Six subjects completed the study. They had a mean (±SD) body mass index of 22.7 ± 1.3 kg/m² (range: 20.9 – 24.3) and a mean (±SD) age of 26.6 ± 3.1 years (range: 23 – 32).

The experiment had a randomized, controlled, single-blinded, crossover design. The duration of the study was 8 days and included a run-in period of three days, two single days of intervention on day 4 and day 8, separated by three days of washout period. During the run-in period on days 1-3 and the wash out period on days 5-7 the volunteers had a controlled diet for breakfast, lunch and dinner. This diet was devoid of dairy products, fermented products and minimized for food with potentially high levels of bioactive nutrients, in particular fruits and vegetables (Table 2). The subjects were advised not to drink more than one cup of coffee a day. Apart from consuming the controlled meals under supervision, no additional control for dietary compliance was made. On day 4 and day 8 after an overnight fast the subjects ingested 540g of either milk or yogurt for breakfast in randomized order. In transcriptomics studies involving humans, particularly in blood cell transcriptome, inter-individual variations are significant and intra-individual variations are relatively small (2, 4). Our study thus, had a crossover design so that each subject can serve as his own control thereby minimizing variations.
Table 2 - Composition of the diet served to the subjects from day 1 to day 8 of the study. ¹Bread and cake were made without yeast and dairy products. ²Snacks were provided every day between 4 and 5 pm.

<table>
<thead>
<tr>
<th>Day</th>
<th>Meal</th>
<th>Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breakfast</td>
<td>White bread toast¹, apricot jam, omelet, carbonated water</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>French fries, turkey schnitzel, puff pastry, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Couscous (wheat semolina), meat sauce, syrup</td>
</tr>
<tr>
<td>2</td>
<td>Breakfast</td>
<td>White bread toast, apricot jam, omelet, carbonated water</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Parboiled rice, lentils, ice sorbet, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Spätzle with meat sauce, cake¹, carbonated water</td>
</tr>
<tr>
<td>3</td>
<td>Breakfast</td>
<td>Potato pancakes, scramble egg, carbonated water, sweet</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Fried grated potatoes, fish finger, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Macaroni with meat sauce, cake, carbonated water</td>
</tr>
<tr>
<td>4</td>
<td>Breakfast</td>
<td>Test dairy product - milk or yogurt</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>French fries, turkey schnitzel, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Couscous with meat sauce, cake, carbonated water</td>
</tr>
<tr>
<td>5</td>
<td>Breakfast</td>
<td>White bread toast, apricot jam, omelet, carbonated water</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>French fries, turkey schnitzel, puff pastry, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Couscous (wheat semolina), meat sauce, syrup</td>
</tr>
<tr>
<td>6</td>
<td>Breakfast</td>
<td>White bread toast, apricot jam, omelet, carbonated water</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Parboiled rice, lentils, ice sorbet, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Spätzle with meat sauce, cake, carbonated water</td>
</tr>
<tr>
<td>7</td>
<td>Breakfast</td>
<td>Potato pancakes, scramble egg, carbonated water, sweet</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Fried grated potatoes, fish finger, syrup</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Macaroni with meat sauce, cake, carbonated water</td>
</tr>
<tr>
<td>8</td>
<td>Breakfast</td>
<td>Test dairy product - yogurt or milk</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>End of the study</td>
</tr>
<tr>
<td>Snacks²</td>
<td></td>
<td>Potatoes chips, cake</td>
</tr>
</tbody>
</table>
Test meals
We used homogenized pasteurized whole milk to prepare two dairy products for the study. One half of the milk was acidified using 2% glucono delta-lactone (GDL) (Jung bunzlauer AG, Basel, Switzerland). GDL is used in dairy industries for controlled acidification in milk and yogurt preparation (62, 63). In this paper acidified milk using GDL is referred to as milk. The other half was used to make conventional yogurt using 3% of starter culture comprising of Lactobacillus delbrueckii subsp. bulgaricus and Streptococcus salivarius subsp. thermophilus. The detailed procedure used for the preparation of milk and yogurt is described in Table 3. Both milk and yogurt were kept similar in weight, volume, energy content, appearance, pH (4.27 for yogurt; 4.12 for milk) to blind the subjects to treatment order. An informal sensory panel composed of six trained persons did not identify significant differences in the texture or the appearance of the two products. A routine microbiological analysis of both products revealed the absence of pathological microbiological contaminants.

Table 3 - Protocol used for the preparation of Yogurt and GDL-milk.

<table>
<thead>
<tr>
<th>Processing</th>
<th>GDL-Milk</th>
<th>Yogurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole milk</td>
<td>97%</td>
<td>97%</td>
</tr>
<tr>
<td>Fat-free powder</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Homogenization 1</td>
<td>65°C / 150 bar</td>
<td>65°C / 150 bar</td>
</tr>
<tr>
<td>Homogenization 2</td>
<td>65°C / 30 bar</td>
<td>65°C / 30 bar</td>
</tr>
<tr>
<td>Pasteurization</td>
<td>92°C / 5 min</td>
<td>92°C / 5 min</td>
</tr>
<tr>
<td>Cooling</td>
<td>30°C</td>
<td>42°C</td>
</tr>
<tr>
<td>GDL&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>Starter culture&lt;sup&gt;2&lt;/sup&gt;</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>Incubation</td>
<td>30°C / up to pH 4.34 / – 7h</td>
<td>42°C / up to pH 4.5 / – 6h</td>
</tr>
<tr>
<td>Cooling</td>
<td>5°C, 1 day</td>
<td>5°C, 1 day</td>
</tr>
<tr>
<td>Storage</td>
<td>&lt; 5°C</td>
<td>&lt; 5°C</td>
</tr>
<tr>
<td>Shelf life</td>
<td>21 days</td>
<td>21 days</td>
</tr>
</tbody>
</table>

The macronutrient composition of milk and yogurt product is described in Table 4.
Table 4 - Macronutrient composition of milk and yogurt. The values for lactose and carbohydrates were not measured and are taken from “Food composition and nutrition tables” by Souci, Fachmann and Kraut, 6th edition, Medpharm scientific publishers. The starter culture was composed of Laactobacillus delbrueckii subsp. bulgaricus and Streptococcus salivarius subsp. thermophilus. GDL, glucono delta-lactone.

<table>
<thead>
<tr>
<th>Composition (per 100 mL)</th>
<th>Milk</th>
<th>Yogurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy (kJ)</td>
<td>284.5</td>
<td>263.6</td>
</tr>
<tr>
<td>Water (mL)</td>
<td>83.2</td>
<td>86.2</td>
</tr>
<tr>
<td>Protein (g)</td>
<td>4.4</td>
<td>4.1</td>
</tr>
<tr>
<td>Fat (g)</td>
<td>3.6</td>
<td>3.3</td>
</tr>
<tr>
<td>Lactose (g)</td>
<td>4.3</td>
<td>3.7</td>
</tr>
<tr>
<td>Other carbohydrates (g)</td>
<td>-</td>
<td>1.1</td>
</tr>
<tr>
<td>GDL (g)</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Starter culture (mL)</td>
<td>-</td>
<td>3</td>
</tr>
</tbody>
</table>

Blood samples

On day 4 and day 8, an intravenous catheter was installed in the forearm of the volunteers by trained nurses to facilitate multiple blood sampling. For gene expression analysis, the first blood samples defined as the 0h sample were drawn immediately prior to the ingestion of the dairy products which was completed within a period of 20-30 minutes. Samples were collected at 2h, 4h and 6h after the ingestion of the dairy products. Blood was collected in triplicate into PAXgene tubes (Preanalytical Systems, Basel, Switzerland) according to the manufacturer's instructions and incubated at room temperature for 9h (64) followed by storing at -80°C until use for further processing. The PAXgene system with RNA stabilizing solution offers several advantages for collection, storage, and processing of whole blood samples in multicentre trials (65). In particular, the isolation of leucocyte subtypes requires immediate processing of the samples on the site of collection, a shortcoming in a multicenter setup such as in our study. Moreover, ex vivo changes in gene expression profiles are induced during the separation of leucocyte subtypes (66). Hence, in order to be able to measure a transcriptional signature that reflects the sum of the changes induced by dairy products; a whole blood analysis was preferred over an analysis restricted to a subset of cells.

In order to identify subjects with sub-clinical signs of infection or inflammation blood samples collected at 0h and 6h on the intervention days were analyzed for C reactive protein (CRP) using immunoturbidimetry method and differential leukocyte count using automatic
differentiation method. Based on this analysis, one subject was excluded from the microarray analysis as he had an abnormal neutrophil count at one 0h time point.

**RNA isolation, labelling, and hybridization**

Prior to RNA isolation, PAXgene tubes were removed from -80°C and allowed to thaw at room temperature overnight. RNA was extracted according to the manufacturer’s instructions. Extracted RNA was stored at -80°C. RNA quality and integrity was verified using the RNA 6000 Nano assay on the Agilent 2100 Bioanalyzer (Agilent Technologies, Basel, Switzerland). RNA concentrations were determined using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Littau, Switzerland). All RNA samples used for microarray analysis had a 260:280 absorbance ratio between 1.9 and 2.1.

Globin mRNA was depleted from a portion of each total RNA sample (1.0μg ± 0.2μg) using the GLOBINclear™-Human kit (Ambion, Huntingdon, UK). A second subject was also excluded from further analysis as he had a RNA yield that was insufficient for microarray hybridization. Globin-depleted RNA samples (200 ng) were amplified using Agilent low RNA input linear amplification kit and control reagents to generate Cy3-labeled antisense cRNA. The labeled cRNA samples were hybridized to Agilent 4x44k Whole Human Genome (G4112F) Oligo Microarray (Agilent Technologies, Basel, Switzerland). The hybridized slides were washed, stabilized, dried, and immediately scanned (Agilent Technologies Microarray Scanner, Agilent Technologies, Switzerland) according to a standard protocol (Agilent Technologies, Basel, Switzerland). Microarray hybridization was performed at the Functional Genomics Center Zurich (FGCZ) microarray core facility, ETH Zurich, Switzerland.

**Data Analysis**

Image processing was performed using Agilent’s *feature extraction* software version 9.5.3.1. The resulting output file contains the intensity of each spot which is summarized by the mean or the median pixel intensity of the spot. The normalization of the median (raw) intensity was carried out using the SNOMAD (standardization and normalization of microarray data) method (67) and R statistical language. Briefly, the SNOMAD approach first performs global normalization and then local mean normalization across each microarray surface to reduce spatial bias if any, followed by log2 transformation of normalized values. The log2 signal ratios (using the 0h time point as the reference) and log2 mean intensities were estimated.
The log2 ratios are normalized across local log2 mean intensities using the ‘loess’ function of R. SNOMAD then finally provides the normalized standardized local Z-score ratio for each analyzed gene. SNOMAD Z-score ratios for the following comparisons were estimated: M2/M0, M4/M0, M6/M0, Y2/Y0, Y4/Y0, Y6/Y0, where M represents milk, Y represents yogurt, and the digits 0, 2, 4, 6 refer to the time points, in hours, at which the blood samples were taken. SNOMAD Z-score ratio comparing gene expression at day 4 and day 8 before the ingestion of milk and yogurt, respectively were also estimated (Y0/M0).

For each of the dairy product, an analysis of variance (ANOVA, SYSTAT version 12.0), more specifically ANOVA repeated measures was performed for each gene using the normalized Z-scores ratio to identify the differentially expressed genes. In addition linear contrasts and quadratic contrasts were tested to identify temporal trends in gene expression over time. In each of the three analyses a stringent estimate of Benjamini-Hochberg false discovery rate (FDR) (68) was obtained using the entire gene dataset. FDR threshold for significance was set at a FDR of 12.5%. Differentially expressed probe sets in the groups of samples from subjects having ingested milk are therefore referred as the “M group” and yogurt as “Y group”. Information on gene annotation and function were retrieved through EntrezGene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db= gene). In order to ascertain that the genes selected as differentially expressed are not due to spatial bias we performed chi-square goodness-of-fit tests using FDR thresholds of 10%, 15% and 20%, respectively, with the null hypothesis of uniform distribution of significant genes across array surface. We found the null hypothesis to be confirmed for all the FDR thresholds for both milk and yogurt datasets.

Spearman rank correlation analysis between gene expression levels in the M and Y groups was performed using R environment. Pathway analysis was performed on M group and Y group by using GENMAPP 2.0 (69) and MAPPfinder (70) to find functional groups of genes (based on Gene Ontology using gene database - Hs-Std_20070817.gdb) that were overrepresented in the subsets. Criteria where z score > 2, permuted p value < 0.05 and more than 3 genes changed in a functional group was used as a threshold for identifying overrepresented functional groups. Additional information on gene function and pathway analysis was retrieved through MetaCore version 5.4 (GeneGo, St. Joseph, MI, USA). The gene expression data in a MIAME format will be made available through ArrayExpress, a public repository for microarray data (http://www.ebi.ac.uk/arrayexpress/).
Gene set enrichment analysis of human genes homolog to bovine proteins present in bovine milk

A list of 183 proteins present in bovine milk were identified according to published references (71, 72). 132 of these have homologues in humans according to the Homologene database. The corresponding gene is present on the Agilent 4x44k Whole Human Genome array G4112F for 112 of them (Additional file 2). Using this set of 112 homologous genes, we first checked, using the gene set enrichment analysis (GSEA) algorithm, whether these 112 proteins were randomly distributed throughout the list of genes for the M group ranked according to the FDR-values of their linear contrasts. In a second analysis we first separated the genes according to the sign of their linear contrasts (positive or negative) before ranking the entire list according to the FDR-values of the linear contrasts. In a separate analysis we also identified differentially expressed human genes in the M and Y groups, e.g. genes with FDR < 12.5% for the linear contrast analysis, that are also present as homologues in the list of 112 bovine proteins.

Gene set enrichment analysis for comparison between the M and Y group

In addition, GSEA was carried out independently on the entire gene list from the M and Y groups to compare and identify pathways differently modulated in these two groups. The genelists of the M and Y groups were ranked based on FDR and linear contrast values. Molecular databases from Biocarta comprising 246 genesets were used for enrichment analysis. An FDR of 5% was used to identify pathways with significant enrichment.
Acknowledgements

This work was supported by Agroscope Liebefeld-Posieux Research Station ALP internal funding. We thank all the volunteers and nurses who took part in the study; Prof. N. Shah (Melbourne, Australia), Prof. I. Johnson (Norwich, UK), Prof. T. Walczyk (Singapore city, Singapore), Prof. M. Zimmerman (Zurich, Switzerland), and Dr. R. Elliott (Norwich, UK) for helpful discussions on the study design; Dr. S. Storcksdieck genannt Bonsmann, Dr. I. Aeberli, Dr. S. Good and Dr. S. Mallia for their assistance during the trial. Dr. D. Thommen and his team (Bern, Switzerland) for technical advice on blood sampling; W. Strahm and C. Hegel (Bern, Switzerland) for the preparation of milk and yogurt; Dr. B. Walther, Dr. R. Sieber, A. Schmid, D. Gille, and M. Chollet at ALP (Bern, Switzerland) for commenting the manuscript; A. Patrignani (FGCZ, Zurich, Switzerland) for his assistance on the microarray experiment, Dr. W. Luginbühl (ChemStat, Bern, Switzerland) and F. Schutz (SIB, Lausanne, Switzerland) for statistical analyses and consultation.
Reference List


CHAPTER 3 – TRANSCRIPTION FACTOR ANALYSIS

Identification of enriched transcription factor binding sites in blood cells in response to milk ingestion in humans

Francina M Sagaya\textsuperscript{a,b}, Richard F Hurrell\textsuperscript{b} and Guy Vergères\textsuperscript{a*}

\textsuperscript{a}Agroscope Liebefeld-Posieux Research Station (ALP), Berne, Switzerland

\textsuperscript{b}Human Nutrition Laboratory, Institute of Food Nutrition and Health, ETH Zurich, Switzerland

\textsuperscript{*}Address correspondence to Dr. Guy Vergères, Biochemistry & Physiology, Agroscope Liebefeld-Posieux Research Station (ALP), Schwarzenburgstrasse 161, 3003 Berne, Switzerland. Tel. +41 (0)31 323 81 54; Fax +41 (0)31 323 82 27; E-mail: guy.vergeres@alp.admin.ch
Abstract

Transcription factors act as biological sensors to environmental signals. As such, transcription factors sense dietary signals and regulate expression of genes that modulate the physiological processes. In this study, we have analysed the promoter regions of the genes that are significantly induced in the postprandial response of blood cells of healthy human subjects to the ingestion of dairy products, in particular milk and yogurt. We identified enriched binding sites for transcription factor belonging to the classes ETS, FORKHEAD, HMG, ZINC-FINGER, HOMEO, MADS, TEA. Interestingly, the transcription factor binding site for GABPA (NRF2), the master regulator of antioxidant response, is selectively enriched in the set of up-regulated genes. Our results provide a holistic picture of key regulatory factors potentially involved in the postprandial response to food ingestion and provide further evidence to the proposed anti-oxidant/anti-inflammatory properties of dairy products.
Introduction

Transcription factors (TFs) are sensors that respond to environmental signals, including dietary signals, by regulating gene expression (1). Various studies have shown that diet and dietary components can directly or indirectly influence the transcription of genes to induce a specific physiological response in the cell (2-4). Hence, there is growing interest in identifying the TFs involved in the response to a diet or to dietary components and to build a genome-wide transcriptional network that could comprehensively describe how these factors act together. In this context, microarray data is now commonly analysed to predict transcription factor binding sites (TFBSs) (5, 6). Indeed, many studies have demonstrated that, by combining computational analysis of cis-regulatory promoter elements with gene expression measurements from microarray data, major TFs that regulate transcriptional networks can be identified, even in complex mammalian systems (7-10). The identification of regulatory transcriptional networks in humans may ultimately help elucidating the mechanisms regulating the transition from health to diseases and may thus participate in the development of preventive therapies (11). Understanding the transcriptional regulation mechanisms activated by the human organism in response to the ingestion of food is therefore of particular interest.

Dairy products, such as milk and yogurt, are important components of Western diets. These products not only provide essential macronutrients, vitamins, trace elements, and minerals but also contain bioactive components that may act synergistically to modulate physiological processes important to the maintenance of health (12). In this context, the impact of dairy products has been investigated in relation to health and the development or the prevention of cancer (13), infection (14), and obesity (15). Significant attention has also been paid to the immunomodulatory properties of dairy products (16, 17).

Typically, gene expression data obtained by microarray analysis can be organized in groups of genes with similar biological functions or with similar expression patterns (i.e. co-expressed genes) with the assumption that co-expressed genes may be co-regulated by a set of TFs. A systematic approach to study how these genes are regulated is to analyze their promoter sequences so as to identify binding motifs for TFs. Computational methods to identify regulatory elements are gaining importance (18) as the experimental approaches used for their identification are slow and laborious, and sometimes even inaccurate (19). In this study, we have chosen the computational strategy and have analysed the 1’200 base pair (bp)
promoter region (1’000bp upstream and 200bp downstream) of the genes that are differentially expressed in blood cells in immediate response to the ingestion of dairy products by humans.
Methods and materials

Study design

Our overall strategy for data analysis is summarized in Figure 1. Briefly, the experiment had a randomized, controlled, single-blinded, crossover design. The entire duration of the study was 8 days and included a three-days run-in period, and two single days of intervention on day 4 and day 8 separated by a three-days washout period. During the run-in and washout periods the subjects were provided with a controlled and restricted diet that was devoid of dairy and fermented products. For the intervention, two dairy products (milk and yogurt) were used and milk was coagulated using glucono-delta lactone (GDL) to keep the study blinded in the crossover design. Six healthy male subjects ingested 540g of GDL-milk or yogurt on day 4 or day 8 of the study period. Blood was collected at 4 time points (0h, 2h, 4h and 6h) so as to evaluate the postprandial effect of milk ingestion using a linear contrast kinetic analysis of the data. Agilent whole human genome array (4x44k (G4112F) Oligo Microarray) was used to identify the genes differentially expressed after the ingestion of the dairy products. The linear contrast analysis revealed 575 and 625 transcripts (FDR 12.5%, \( p \leq 0.002 \)) that were differentially expressed after the ingestion of milk and yogurt by healthy subjects, respectively (Sagaya et al, manuscript submitted). The data has been deposited in ArrayExpress and will be made available from May 2011. In the present study, we used genes which showed differential expression after the ingestion of milk and yogurt and are referred to as M group and Y group, respectively.
Figure 1 Illustrates the study design and gives an overview of the workflow involved in the transcription factor enrichment analysis. *milk coagulated using Glucono-delta lactone (GDL).

**Input data for promoter analysis**

A common method to analyze regulatory relationships among genes using microarray data is to cluster the genes, based on their expression profiles or biological function, into sets of putatively co-regulated genes (20). For the extraction of promoter sequences, we separated the differentially expressed genes into up-regulated and down-regulated genes, based on their linear score (e.g. either positive or negative values, respectively). Using the Toucan version 3.1.0 software we then extracted the promoter sequences covering 1’000 bp in the upstream region and 200 bp in the downstream region of the annotated transcription start site for each gene in the target set of genes (21). The setting used to extract the promoter regions from Ensembl database (22) was as follows: get-sequence, ID file, Human, Agilent probes, region between upstream 1’000 bp and downstream 200 bp of 5’ upstream exon 1, reverse complement automatically.

**Enrichment analysis of TFBSs**
Three sets of promoter sequences were extracted for motif enrichment analysis of the differentially expressed genes. The three datasets used were 1) 302 up-regulated genes and 128 down-regulated genes in M group, 2) 306 up-regulated genes and 181 down-regulated genes in Y group, and 3) differentially expressed genes belonging to specific functional categories enriched in Gene Ontology (GO), in particular the biological processes ‘translation’, ‘apoptosis’, ‘immune response’, and ‘inflammatory response’ and the cellular component ‘mitochondrial part’. A library-based motif enrichment analysis was carried out to characterize the transcriptional regulatory factors regulating the transcriptional response observed in the human postprandial study. This type of analysis compares a set of known TFBS to the regulatory regions of a set of co-regulated genes, and uses statistical tests to identify those regulatory motifs that are found more often - or less often - than would be expected by chance alone. We employed the Clover software (23) to screen the JASPAR CORE database (2005) (24), which contains 123 TFBS. We set the Clover parameters to 2’000 randomizations and a p-value threshold of 0.05. The p-values were estimated by separately comparing the scores obtained with our gene sets to three different background sets: 1) chromosome 20 sequences (7 sequences, 59’424’990 bp, 44.1% CG content), 2) CpG Island (27’555 sequences, 20’891’518 bp, 68.8% CG content), and 3) 2’000 bp upstream of human promoter sequences (17’516 sequences, 35’032’000 bp, 49.8% CG content). The content of the background sets determines which motifs will be identified as significantly over-represented in the sequences of interest. The CG content of our sequences of interest in the 1’200 bp region investigated was > 50 % which is comparable to the background sequence sets and in line with the expected values for the CG content of the promoter region of genes (8).

For each dataset, we report the TFBSs enriched at p < 0.05 in all three statistical tests (e.g. with all three background sets). The STAMP tool-kit was further used to match JASPAR and TRANSFAC matrices using default parameters (25).
Chapter 3

TRANSCRIPTION FACTOR ANALYSIS

Result and discussion

The computational identification of TFBSs is key in understanding gene regulation but these methods can also generate significant numbers of false positives. In this analysis, we have employed stringent criteria to minimize false positives. Firstly, we used the JASPAR database which represents a curated and non-redundant dataset (24). Compared to other databases (e.g. TRANSFAC), JASPAR motifs are high quality binding sites that are derived exclusively from sets of nucleotide sequences experimentally demonstrated to bind transcription factors (26). Secondly, for each dataset investigated, we selected only those motifs enriched at p < 0.05 in all three statistical tests thereby minimizing the chances of finding false positives in our set of significantly enriched motifs. Table 1 shows that, out of the 123 JASPAR motifs analyzed, 18 motifs (15%) were enriched in the 1’200bp region of the up-regulated and down-regulated genes in the M group.

Table 1 Occurrence of enriched JASPAR motifs in the sets of genes up- and down-regulated in M and Y groups.

<table>
<thead>
<tr>
<th>JASPAR motif</th>
<th>Class</th>
<th>Milk-up</th>
<th>Milk-down</th>
<th>Yogurt-up</th>
<th>Yogurt-down</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA0003 TFAP2A</td>
<td>AP2</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0123 ABI4</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0049 Hunchback</td>
<td>C2H2-ZN-Finger</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0060 NF-Y</td>
<td>CAAT-box</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>MA0062 GABPA</td>
<td>ETS</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>MA0076 ELK4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>MA0033 FOXL1</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0040 Foxq1</td>
<td>FORKHEAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0041 Foxd3</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0042 FoxI1</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0045 HMG-JY</td>
<td>HMG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0084 SRY</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0075 Prrx2</td>
<td>HOMEO</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MA0008 Athb-1</td>
<td>HOMEO-ZIP</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MA0082 SQUA</td>
<td>MADS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0090 TEAD</td>
<td>TEA</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0100 Broad-complex_1</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MA0011 Broad-complex_2</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0012 Broad-complex_3</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0013 Broad-complex_4</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+ indicates the occurrence of the motif in the mentioned category.
The raw scores and the associated p-values for each of these motifs are shown in Table 2. The enriched motifs were grouped into 10, out of 29, TF classes and comprised AP2, C2H2-Zn FINGER, CAAT-BOX, ETS, FORKHEAD, HMG, HOMEQ, HOMEQ-ZIP, MADS, TEA, and ZN-FINGER. In 5 out of 10 TF classes, and in 7 out of 18 TFBSs, the binding sites were shared for both the up-regulated and the down-regulated sets of genes (Table 1).

**Table 2** Significant JASPAR motifs in sequences upstream of up- and down-regulated genes in M group. A p-value of 0.00 denotes a value <0.010

<table>
<thead>
<tr>
<th>Motif</th>
<th>Name</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcription factors over-represented in Up-regulated genes in M group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0045*</td>
<td>HMG-IY (HMG)</td>
<td>443</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Lymphocyte differentiation [28]</td>
</tr>
<tr>
<td>MA0010*</td>
<td>Broad-complex_1 (ZN-FINGER, C2H2)</td>
<td>324</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0013*</td>
<td>Broad-complex_4 (ZN-FINGER, C2H2)</td>
<td>248</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0049*</td>
<td>Hunchback (ZN-FINGER, C2H2)</td>
<td>305</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Closest sequence homology to Ikaros which is involved in Immune development [33]</td>
</tr>
<tr>
<td>MA0082*</td>
<td>SQUA (MADS)</td>
<td>164</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>MA0041*</td>
<td>Fox3 (FORKHEAD)</td>
<td>255</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Maintenance of neural crest and pluripotent cells (Pandey, Munjal &amp; Datta, 2010)</td>
</tr>
<tr>
<td>MA0042*</td>
<td>Fox1 (FORKHEAD)</td>
<td>116</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>Cell development and organogenesis [32]</td>
</tr>
<tr>
<td>MA0003</td>
<td>TFAP2A (AP2)</td>
<td>77</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>Immune response</td>
</tr>
<tr>
<td>MA0062</td>
<td>GABPA (ETS)</td>
<td>31</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>Stress response [41]</td>
</tr>
<tr>
<td><strong>Transcription factors over-represented in Down-regulated genes in Y group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0045*</td>
<td>HMG-IY (HMG)</td>
<td>481</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Lymphocyte differentiation [28]</td>
</tr>
<tr>
<td>MA0084</td>
<td>SRY (HMG)</td>
<td>116</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>transcriptional activator in lymphocytes</td>
</tr>
<tr>
<td>MA0010*</td>
<td>Broad-complex_1 (ZN-FINGER, C2H2)</td>
<td>344</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0013*</td>
<td>Broad-complex_4 (ZN-FINGER, C2H2)</td>
<td>287</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0012</td>
<td>Broad-complex_3 (ZN-FINGER, C2H2)</td>
<td>190</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0011</td>
<td>Broad-complex_2 (ZN-FINGER, C2H2)</td>
<td>87</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Regulates apoptosis in Drosophila [34]</td>
</tr>
<tr>
<td>MA0049*</td>
<td>Hunchback (ZN-FINGER, C2H2)</td>
<td>309</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Immune development [33]</td>
</tr>
<tr>
<td>MA0082*</td>
<td>SQUA (MADS)</td>
<td>207</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>MA0041*</td>
<td>Fox3 (FORKHEAD)</td>
<td>260</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Maintenance of neural crest and pluripotent cells (Pandey et al., 2010)</td>
</tr>
<tr>
<td>MA0042*</td>
<td>Fox1 (FORKHEAD)</td>
<td>130</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>cell development and organogenesis [32]</td>
</tr>
<tr>
<td>MA0040</td>
<td>Foxq1 (FORKHEAD)</td>
<td>45</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>Promotes/modulates natural killer cell activity [20]</td>
</tr>
<tr>
<td>MA0033</td>
<td>FoxL1 (FORKHEAD)</td>
<td>45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>maintenance of gut-associated lymphoid organs [32]</td>
</tr>
<tr>
<td>MA0123</td>
<td>ABI4 (AP2)</td>
<td>77</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>MA0090</td>
<td>TEAD (TEA)</td>
<td>15</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>MA0075</td>
<td>Prx2 (HOMEQ)</td>
<td>23</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>MA0008</td>
<td>AthB-1 (HOMEQ-ZIP)</td>
<td>39</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Motifs present in both up and down-regulated gene list are indicated with an asterisk.

-A- Raw score; B- p-value relative to Human chromosome 20; C- p-value relative to Human CpG islands

D- p-value relative to Human promoters

The presence of some of these enriched motifs in both the up-regulated and down-regulated sets of genes could be in line with previous studies showing that TFs may act both as transcriptional activators or repressors based on their microenvironment (28). For instance,
HMG-I/Y can either activate or repress the expression of different immune related genes (29). However, the biological mechanisms that turn the activity of a TF from an activator to a repressor remain unclear (30). A clue to this phenomena may be found in Drosophila transcription factor Hunchback whose activity has been postulated to be modulated by its own cellular concentration (30).

The functional role of most of the enriched TFs in immune cells is not fully understood. Members of the FORKHEAD family are gaining importance for their role in immune functions (31). For example, Foxq1 is an effector of the function of natural killer cells and activation of FoxL1 helps maintaining the function of the gut-lymphoid organ system (32, 33). In Drosophila, Hunchback proteins (that have close sequence homology to the human Ikaros proteins) have been associated with immune development (34). Also, the Broad-complex proteins have been found to regulate apoptosis (35). The presence of these enriched motifs in human immune cells after the ingestion of dairy products suggest the importance of immune-related functions in postprandial metabolism (see Table 1).

As milk and yogurt are similar in their composition and in their nutritional properties a similar gene expression pattern may reasonably be expected for both datasets. Our previous study (Sagaya et al, manuscript submitted) revealed a similar physiological response of human subjects to the ingestion of milk and yogurt. Indeed, even though only 11% of the differentially expressed genes overlapped between the two groups, a broader analysis of the linear contrast scores revealed a significant correlation between the two sets of differentially expressed genes. In order to get an holistic view of the regulatory mechanisms involved in the postprandial response, this investigation compared the TFBS identified in the set of genes up-regulated (306) as well as down-regulated (181) in Y group with that of M group. The promoter sequences of the genes in Y group showed an enrichment of 12 motifs out of the 123 motifs present in the JASPAR data bank (10%). The Y group also showed a similar enrichment in the classes of TFs AP2, C2H2-ZN-FINGER, ETS, FORKHEAD, HMG, MADS, and ZN-FINGER (Table 1). Out of 11 motifs identified in the up-regulated set of genes, 8 were found in both the M and Y groups. For the down-regulated set of genes, all the enriched motifs identified in the Y group were also enriched in the M group (Figure 2). This indicates that yogurt tend to globally induce a postprandial transcriptional response similar to that of milk in humans.
Figure 2 Venn diagram showing the enriched motifs in milk and yogurt datasets. Almost all enriched motifs in down-regulated genes in the yogurt dataset are shared by the milk dataset motif whereas 3 motifs in up-regulating genes are unique to the yogurt dataset.

Genes may show similar expression profiles for reasons other than co-regulation based on gene expression pattern (20). As genes involved in the same biological processes or pathways are more likely to be co-regulated (36), grouping genes based on their function prior to conducting a TFBS analysis may thus increase the accuracy in predicting TFBS (37). Indeed, many known cis-regulatory elements have been successfully recovered by applying promoter motif search tools to functional groups of genes (38, 39). As many of the enriched motifs for the M and Y groups were similar, we next focused our analysis on the identification of motifs in subsets of genes in these two groups that possess common biological functions.

We have previously identified a set of biological processes based on GO category commonly modulated as a result of ingestion of milk or yogurt (Sagaya et al, manuscript submitted). In particular, genes were up-regulated in the processes referred to as ‘translation’ (biological process) and ‘mitochondria’ (cellular component) and down-regulated in the biological processes ‘apoptosis’, immune response’, and ‘inflammation’. In the present study, we have identified statistically over-represented TFBS in the promoter regions of the genes belonging to these functional groups. Compared to the overall list of genes (Milk up, Milk down, Yogurt up, Yogurt down) motif enrichment in functional groups revealed more specific TFs (Tables 3 and 4).
Table 3 Significant JASPAR motifs in sequences upstream of up- and down-regulated functional categories in M group.

<table>
<thead>
<tr>
<th>Motif</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>One motif occurrence/sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation (37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA (NRF2)</td>
<td>10</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>28/37</td>
</tr>
<tr>
<td>HMG-1Y</td>
<td>73</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>34/37</td>
</tr>
<tr>
<td>Hunchback</td>
<td>55</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>33/37</td>
</tr>
<tr>
<td>SQUA</td>
<td>42</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>33/37</td>
</tr>
<tr>
<td>Foxd3</td>
<td>41</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>33/37</td>
</tr>
<tr>
<td><strong>Mitochondrial part (24)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA (NRF2)</td>
<td>5</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>17/24</td>
</tr>
<tr>
<td><strong>Down-regulated process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apoptosis (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foxd3</td>
<td>29</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>15/17</td>
</tr>
<tr>
<td>SQUA</td>
<td>18</td>
<td>0.00</td>
<td>0.05</td>
<td>0.01</td>
<td>16/17</td>
</tr>
<tr>
<td>Foxx1</td>
<td>17</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>15/17</td>
</tr>
<tr>
<td>Prrx2</td>
<td>3</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>14/17</td>
</tr>
<tr>
<td>ATHB5</td>
<td>0.5</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>14/17</td>
</tr>
<tr>
<td><strong>Immune response (17)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athb-1</td>
<td>7</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>12/17</td>
</tr>
<tr>
<td>IRF1</td>
<td>17</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>11/17</td>
</tr>
<tr>
<td>NF-κB</td>
<td>4</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>8/177</td>
</tr>
<tr>
<td>Pbx</td>
<td>2</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>14/17</td>
</tr>
<tr>
<td>RELA</td>
<td>5</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>14/17</td>
</tr>
<tr>
<td><strong>Inflammatory response (11)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10- USP</td>
<td>2</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>8/11</td>
</tr>
<tr>
<td>NF-κB</td>
<td>3</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
<td>8/11</td>
</tr>
<tr>
<td>Fos bZIP</td>
<td>4</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>9/11</td>
</tr>
</tbody>
</table>

*A- Raw score; B- *p*-value relative to Human chromosome 20; C- *p*-value relative to Human CpG islands

D- *p*-value relative to Human promoters Numbers within the parenthesis indicates number of genes in the mentioned category.
Table 4 Significant JASPAR motifs in sequences upstream of up- and down-regulated functional categories in Y group.

<table>
<thead>
<tr>
<th>Motif</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>One motif occurrences/sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation (37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA (NRF2)</td>
<td>23</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>28/37</td>
</tr>
<tr>
<td>E74A</td>
<td>18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>33/37</td>
</tr>
<tr>
<td>ELK1</td>
<td>2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>18/37</td>
</tr>
<tr>
<td>SU-h</td>
<td>7</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>26/37</td>
</tr>
<tr>
<td><strong>Mitochondrial part (11)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ar</td>
<td>2</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>6/11</td>
</tr>
<tr>
<td>ELK1</td>
<td>4</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>10/11</td>
</tr>
<tr>
<td>NF-Y</td>
<td>7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9/11</td>
</tr>
<tr>
<td>ABH4</td>
<td>5</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>11/11</td>
</tr>
<tr>
<td><strong>Down-regulated process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apoptosis (16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hunchback</td>
<td>27</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>NF-κB</td>
<td>3</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>15/16</td>
</tr>
<tr>
<td>SRF</td>
<td>0.4</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>REL</td>
<td>6</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>12/16</td>
</tr>
<tr>
<td>Spz1</td>
<td>5</td>
<td>0.04</td>
<td>0.00</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td><strong>Immune response (11)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF2-II</td>
<td>4</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>3/11</td>
</tr>
</tbody>
</table>

* A- Raw score; B- p-value relative to Human chromosome 20; C- p-value relative to Human CpG islands D- p-value relative to Human promoters Numbers within the parenthesis indicates number of genes in the mentioned category.

In particular, the up-regulated process ‘translation’ showed a significant enrichment of the motif GABPA (also referred to as NRF2) in both the M and Y groups, whereas the down-regulated functional processes “immune response”, “inflammatory response”, and ‘apoptosis’ showed an enrichment for IRF1, RELA, NF-κB binding motifs.

NRF2 activates the expression of anti-oxidant genes. Interestingly, we observed an up-regulation of anti-oxidant genes such as superoxide dismutase 1, thioredoxin 2, selenoprotein H, selenoprotein M and selenoprotein W (p < 0.05) suggesting that the ingestion of milk and yogurt may enhance a postprandial anti-oxidant response via transcriptional regulation of genes under the control of NRF2. This conclusion is in line with previous study proposing that human milk enhance anti-oxidant defenses, in particular to eliminate reactive oxygen species (ROS) (40). In addition, bovine milk also contains soluble anti-oxidants which may contribute to anti-oxidant response (41).

An intriguing observation is that most of the genes containing the NRF2 motif in our set of postprandial genes are involved in the translation process and, in particular, code for...
ribosomal proteins. Increasing evidence suggests that a rapid onset of selective protein translation serves as an important mechanism for cells to deal with cellular stress such as oxidative and inflammatory stress (42). For example, a rapid increase in NRF2 protein levels takes place in cardiomyocytes exposed to oxidants (43). Furthermore, recent work concludes that some of the ribosomal proteins which form an important part of the translational machinery contain NRF2 binding sites (44) suggesting that this TF may have a role as a regulator of the expression of ribosomal genes during the stress response. In addition, there is a growing evidence suggesting that ribosomal proteins may have dual function (45). We, thus, postulate that NRF2 may coordinate the expression of ribosomal genes that might be important for orchestrating the anti-oxidative response. Finally, in addition to its role in the antioxidant response, an anti-inflammatory activity for the NRF2 protein has also been reported (46).

Crosstalk between oxidative stress and inflammation is well documented. Among the transcription factors induced by oxidative stress, in particular by ROS, NF-κB plays a crucial role by regulating the expression of genes involved in inflammation and apoptosis. Some nutrients with anti-oxidative activity can inhibit pathways activated by NF-κB as well as enhance the activity of pathways activated by NRF2 (47). Also, NRF2 prevents the up-regulation of NF-κB activity (48). In our study, we identify NF-κB and AP-1 (Fos bZIP), two TFs that are known to be activated by ROS.

In conclusion, our results provide a holistic picture of key regulatory factors (TFBS for TFs NRF2, FORKHEAD, HMG, Broad-complex) potentially involved in the postprandial response to dairy ingestion. In addition, the set of up-regulated genes involved in translation and mitochondrial activity are enriched in the NRF2 motif and the sets of down-regulated genes involved in inflammation and immuno-modulatory functions are enriched in the NF-κB motif. These observations could thus be indicative of an anti-oxidative and anti-inflammatory postprandial response of the human organism to dairy products. On a broader nutritional basis, the postprandial regulation of NF-κB and NRF2 activities in immune cells in response to the ingestion of food may therefore be an interesting target for research and development under conditions in which an elevated level of oxidative stress and associated inflammation predominate.
ACKNOWLEDGEMENTS

This work was partly supported by Agroscope Liebefeld-Posieux Research Station ALP internal funding and partly by international milk genomics consortium (IMGC). The author's responsibilities were as follows - FMS: experimental part of the study, in particular the CLOVER analysis, writing of the manuscript; FMS, RFH and GV: participated in critical revising of the manuscript; none of the authors has a personal or financial conflict of interest. Nutrition group members at ALP (Dr. B. Walther, Dr. R. Sieber, A. Schmid, and D. Gille and M. Chollet (Bern, Switzerland) for their feedback; Dr. Frédéric Shutz (Swiss institute for bioinformatics SIB, Lausanne, Switzerland) for consultation.
References


CHAPTER 4 – MOUSE STUDY

Blood cell transcriptomics: a diagnostic tool for assessing the infection of mice with enterohemorrhagic *Escherichia coli* O157:H7 and the antimicrobial properties of *Lactobacillus gasseri* K7(Rif*)

Francina M. Sagaya¹,², Biljana Hacin³, Gorazd Tompa⁴, Alojz Ihan⁵, Špela Škrajnar⁶, Manica Černe⁶, Bojana B. Matijašič⁶, Irena Rogelj⁶, Richard F. Hurrell² and Guy Vergères¹*

¹Agroscope Liebefeld-Posieux Research Station (ALP), Berne, Switzerland

²Human Nutrition Laboratory, Institute of Food Nutrition and Health, ETH Zurich, Switzerland

³Veterinary Faculty, University of Ljubljana, Slovenia

⁴Biotechnical Faculty, University of Ljubljana, Slovenia

⁵Medical Faculty, University of Ljubljana, Slovenia

⁶Lek Pharmaceuticals d.d., Ljubljana, Slovenia

*Corresponding author. E-mail: guy.vergeres@alp.admin.ch; Fax - +41 (0)31 323 82 27; Tel. +41 (0)31 323 81 54
Abstract

Pathogenic and commensal bacteria colonize the gastrointestinal tract of living organisms by adhering to the gut epithelium and activating underlying immune cells. Measuring gene expression in circulatory immune cells may thus be a powerful approach to study interactions between these bacteria and the host organism. To this end, we have measured the blood cell transcriptome of five groups of mice: (1) control C57BL/6J mice; (2) mice inoculated with enterohemorrhagic *Escherichia coli* O157:H7; (3) mice inoculated with *Lactobacillus gasseri* K7(Rif\(^\text{r}\)); (4) mice co-inoculated with *E. coli* O157:H7 and *L. gasseri* K7(Rif\(^\text{r}\)); (5) mice pre-inoculated with *L. gasseri* K7(Rif\(^\text{r}\)) and subsequently inoculated with *E. coli* O157:H7.

No clinical signs of infection in the *E. coli* O157:H7 group of mice were observed. However, the blood cell transcriptome could clearly distinguish between the five treatment groups. In the *E. coli* O157:H7 group 1’541 genes were differentially expressed and a functional analysis revealed changes in the expression of genes that regulate pathways characteristic of bacterial infection (cell adhesion, cytoskeleton rearrangement, inflammation, glycolysis). Furthermore, inoculation with *L. gasseri* K7(Rif\(^\text{r}\)) suppressed the gene expression profile induced by *E. coli* O157:H7, the effect being more pronounced in the pre-incubation than in the co-incubation protocol.

The blood cell transcriptome thus provides a diagnostic tool that can be employed to assess early molecular events leading to infection as well as renders valuable information, in otherwise asymptomatic situations, on the anti-microbial potential of probiotic bacterial strains.

Keywords: *Escherichia coli* O157:H7, *Lactobacillus gasseri* K7(Rif\(^\text{r}\)), probiotic, transcriptomics
Introduction

Gut epithelial cells are the interface between adhering bacteria and underlying immune cells (6, 18, 35). Biological crosstalk between adhering bacteria on the apical side of the gut and the immune cells on the basolateral side is a key characteristic of this dynamic interface (8, 9, 16). The nature of this crosstalk strongly depends on the type of host-cell interaction, as pathogens (19, 42) evidently modulate the immune response differently than commensal bacteria (23). Characterizing the immune response is therefore a key task in order to understand the mechanisms of action employed by bacteria entering and colonizing the organism via the gastro-intestinal tract.

*E. coli* O157:H7 is an enterohemorrhagic Shiga toxin producing bacteria that causes intestinal inflammation, severe bloody diarrhea, and hemolytic-uremic syndrome (HUS) (17). Infection by this important food-borne pathogen is a problem worldwide, especially in developing nations where it is a cause for mortality, particularly among children. Much of the pathogenicity associated with *E. coli* O157:H7 is attributed to the production of Shiga toxin as it is responsible for HUS. A strong inflammatory response is observed during *E. coli* O157:H7 infection (1, 15, 31) and cytokines produced by the infected macrophages contribute to the severe inflammation associated with HUS (24). Specific treatment and preventive measures for *E. coli* O157:H7 infections are still being explored. Moreover, antibiotic treatment may increase the severity of infection (13). Research is therefore needed to develop alternative therapies for preventing or limiting infections with *E. coli* O157:H7.

Probiotics are live non-pathogenic microorganisms that confer health benefits to the host (34). Many strains of probiotics possess antimicrobial properties and there is a growing interest in employing these strains as therapeutics for reducing the severity of *E. coli* O157:H7 infections. Multiple studies have demonstrated the anti-bacterial activity of *Lactobacillus* strains against gastro-intestinal pathogens (4, 40, 46, 48). In that context, *L. gasseri* K7 has been previously shown to limit the severity of *E. coli* infection *in vitro* on Caco-2 cells and *in vivo* in gnotobiotic pigs (27, 29, 43). This anti-microbial activity was attributed to the modulation of immune response, production of organic acids and bacteriocins, and the competitive exclusion of *E. coli* from intestinal mucosa (43).

In order to develop effective therapeutic protocols there is a need to better understand the molecular events leading to the pathogenesis of *E. coli* O157:H7 as well as to confirm the
anti-bacterial and immunomodulatory properties of *L. gasseri* K7. In particular, the interactions between *E. coli* O157:H7 and immune cells are important in the global pathogenesis of *E. coli* O157:H7 (36, 38) but this interactions are to this end still poorly characterized.

Peripheral blood provides an attractive surrogate tissue for the discovery of biomarkers of hematological and non-hematological diseases (41). In particular, white blood cells are key mediators of the immune response and, as such, an important source of biomarkers to characterize the interactions between the host and pathogenic or commensal bacteria. Gene expression profiling of blood cells has thus a large potential as a diagnostic and research tool in clinical microbiology, in particular in the fields of infectious diseases and probiotic research (8, 42).

The aim of this study was to investigate the potential of blood cell transcriptomics to characterize the infection of C57BL/6J mice with *E. coli* O157:H7 as well as to substantiate the antimicrobial properties of *L. gasseri* K7 in a co-inoculation mouse model.

**MATERIALS AND METHODS**

**Bacterial strains and culture conditions.** *L. gasseri* K7 is an isolate from baby's faeces that was previously identified as a potential probiotic (28) and deposited in a culture collection at the Chair of Dairy Science, Biotechnical Faculty, University of Ljubljana (IM 105) and in the Czech Collection of Microorganisms (CCM 7710). A derivative of *L. gasseri* K7 strain, *L. gasseri* K7(Rif<sup>®</sup>), resistant to rifampicin (250 µg/ml<sup>-1</sup>) was obtained by sub-culturing *L. gasseri* K7 cells in MRS broth (Merck, D-64271 Darmstadt, Germany) with increasing concentration of added rifampicin (Sigma-Aldrich Chemie, D-89552 Steinheim, Germany) at 37°C for 18 hours (26). A clinical isolate of toxigenic *E. coli* O157:H7 was kindly provided by the Laboratory for Bacteriological Diagnostics of Intestinal Infections, Institute of Microbiology and Immunology, Medical faculty, University of Ljubljana. *L. gasseri* K7(Rif<sup>®</sup>) was cultured at 37°C in MRS broth (Merck, Germany) in a microaerophilic atmosphere obtained with the GenBox system (BioMerieux, France). *E. coli* O157:H7 was cultured aerobically at 37°C in BHI broth (Merck, Germany).

**Mouse colonization experiments.** Female C57BL/6J mice were obtained at eight weeks of age from the Medical Experimental Centre (Institute of Pathology, Medical Faculty,
University of Ljubljana). Mice were housed at the Centre for Experimental Animals (Institute of Microbiology and Immunology, Medical Faculty, University of Ljubljana) in a temperature controlled environment with a 12h light-dark cycle. Animal experiments were approved by the ethical committee of the Veterinary Administration of the Republic of Slovenia (license number 34401-29/2008/7) and all treatments were performed in accordance with the principles and procedures outlined in the European Guidelines for Care and Use of Experimental Animals. Mice were fed sterilized solid rodent chow and water. Each mouse was assessed daily for weight gain, feed intake, clinical status and behavioral changes.

Feeding-infection experiments. Mice were randomly divided into five groups containing ten animals each. The study duration (Fig. 1) was 28 days and involved three phases, namely an adaptation phase (days 1–7), a treatment phase (days 8-21), and a washout period (days 22-28). During the 14-days treatment period each mouse in the *L. gasseri* K7(Rif*) group, the pre-incubation group and the co-incubation group received a daily dose of 1x10⁹ viable *L. gasseri* K7(Rif*) cells, determined as colony forming units (CFU) on MRS agar (Merck, Germany), and re-suspended in 100 µl of saline solution. The suspension was applied orally with plastic feeding tubes for rodent oral gavage (Instech Solomon, Plymouth Meeting, PA, USA). A single dose of *E. coli* O157:H7 was applied on day 8 (*E. coli* group, co-incubation group) or on day 15 (pre-incubation group) of trial. Each mouse received a dose of 5x10⁸ viable *E. coli* O157:H7 cells, determined as CFU on *E. coli* O157:H7 agar (Merck, Germany), resuspended in 100 µl of saline solution and introduced orally with plastic feeding tubes. Animals in the control group received 100 µl of saline solution orally during the treatment phase.
**Figure 1.** Study design showing the three phases of the experimental period namely the adaptation phase (7 days), the treatment phase (14 days) and the washout phase (7 days). Five groups of 10 mice underwent the indicated treatments. Blood samples obtained from 4 mice from each group were used for blood cell transcriptomic analysis.

**Blood collection.** Mice were euthanized with CO₂ and blood samples of approximately 600 μl were obtained from each mouse by cardiac puncture. One hundred microliters of blood were collected in microtainer tubes containing EDTA (Microtainer Brand Tubes with EDTA, Becton Dickinson, Franklin Lakes, NJ) to assess hematology parameters. Samples were mixed gently and put on a roller mixer until analysis within 1h after collection. For the phagocytosis assay 100 μl of blood were collected in microtainer tubes containing heparin (Microtainer Brand Tubes with Heparin, Becton Dickinson, Franklin Lakes, NJ). The tubes were immediately placed on ice before further analysis. For microarray analysis 300 μl of blood from four mice per group were transferred immediately after the cardiac puncture to tubes containing a pre-loaded RNA-later solution (Mouse Ribopure, Ambion, Huntington, UK). The samples were stored at -20°C until further processing.
Hematology analysis. Blood from five animals per group was analyzed using an ABC VET hematology analyser (ABX Diagnostics, Montpelier, France) in which the counting of blood cells is based on the volumetric impedance method, directly measuring white blood cells (WBC), red blood cells (RBC), hemoglobin (HGB), hematocrit (HCT), platelets (PLT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH) and MCH concentration (MCHC).

Manual determination of white blood cell differentiation count. Blood smears were made from a drop of EDTA-treated, air-dried blood, and stained (Hemacolor Stain for Microscopy, Merck, Darmstadt, Germany). The slides were examined under light microscopy (Olympus BX51, Europa Holding Company, Hamburg, Germany), and 150-200 white blood cells from each mouse were counted. To eliminate inefficiencies associated with mechanical desktop tally counters, we used a counting software (EasyCell Counter). Each leukocyte was assigned to one of the following categories: lymphocytes, monocytes, segmented neutrophils, band neutrophils, eosinophils, basophils, and other cells. For each category, the percentage was calculated by dividing the number of cells in each category of leukocytes by the total number of leukocytes counted.

Phagocytosis. Phagocytosis was assessed using a test kit for the quantification of phagocytic activity of monocytes and granulocytes in heparinized whole blood (Phagotest, ORPEGEN Pharma, Heidelberg, Germany) according to the manufacturer’s instructions. In brief, 100 µl heparinized whole blood was mixed with 20 µl of a FITC-labeled E. coli bacterial strain (1x10⁹ bacteria per ml) and incubated for 10 min. at 37°C in a water bath. A negative control remained in the ice bath for 10 minutes at 4°C. FITC-positive and FITC-negative phagocytic cells were determined with the use of FACSCalibur flow cytometer (Becton Dickinson, USA) using the blue-green excitation light (488 nm argon-ion laser).

Statistical analysis of data from hematological and phagocytotic tests was performed using an unpaired two-tailed Student’s t-test. Values of p < 0.05 were considered as statistically significant.

Preparation of cDNA for gene expression analysis. For RNA isolation, the tubes containing a pre-loaded RNA-later solution were removed from their storage at −20°C and allowed to thaw at room temperature for 30 min. RNA was extracted according to the manufacturer’s instructions (Ambion, Huntington, UK) and stored at -80°C. The RNA
quantity and integrity was measured at 260 nm in a NanoDrop ND 1000 spectrophotometer (NanoDrop Technologies, Delaware, USA) and electrophoretically in an Agilent 2100 Bioanalyzer using the Eukaryote Total RNA Nano assay (Agilent, Basel, Switzerland). For microarray analysis, high quality RNA was sent to a microarray facility (Functional Genomics Center Zurich, ETHZ, Switzerland). The cDNA generation, amplification and labeling was carried out using the WT-Ovation Pico System according to manufacturer’s instructions (NuGEN Technologies, Inc., San Carlos, USA). Briefly the cDNA was prepared from total RNA using a primer mix and reverse transcriptase (WT-Ovation Pico System). The primers contain DNA sequences that hybridize either to the 5’ portion of the poly (A) sequence or randomly across the transcripts. A linear isothermal DNA amplification process (SPIA amplification) was used to prepare single-stranded cDNA in the antisense direction of the mRNA starting material. The quality and quantity of single-stranded cDNA was determined spectrophotometrically using NanoDrop ND 1000 and Agilent 2100 Bioanalyzer as described above. Single-stranded cDNA (3 μg) was converted into sense target cDNA using the Ovation RNA Amplification System V2 and treated with Ovation WB reagent to increase the cDNA yield. Fragmented and biotin-labeled sense target cDNA was generated with the FL-Ovation cDNA Biotin Module V2 system according to the manufacturer’s instructions (NuGEN Technologies, Inc., San Carlos, USA).

Array hybridization and gene expression analysis. Biotin-labeled single-stranded cDNA (5 μg) from 20 samples (4 samples from each group of mice) were mixed in 220 μl of hybridization mix containing hybridization controls. The samples were hybridized separately onto mouse microarrays for 18h at 45°C (GeneChip Mouse Genome 430 2.0, Affymetrix Inc., Santa Clara, USA). The arrays were then washed in a fluidics station (Fluidics Station 450 FS450 0001, Affymetrix Inc.) according to the manufacturer’s protocol. The fluorescent intensity emitted by the labeled target was then scanned (GeneChip Scanner 3000, Affymetrix Inc.). Raw data processing was performed using the AGCC software (Affymetrix Inc.). After hybridization and scanning, probe cell intensities were calculated and summarized for the respective probe sets using the RMA algorithm (21). MAS5 algorithm was also applied for determining the “present calls” in each array. All arrays showed good quality according to established quality control criteria from Affymetrix protocol. RMA normalized values on 45’101 probesets (21’248 based on Entrez GeneID) were further filtered based on the criteria that (1) the probeset signal intensity is larger than 25 (log2 > 4.6)
and (2) the MAS5 “present call” is true for each probeset in at least one sample (out of 20). The filtered probesets contained 22’855 probes (12’138 genes based on Entrez GeneID).

To identify genes with statistically significant changes in gene expression between control group and treated groups (E. coli O157:H7, pre-incubation, co-incubation and L. gasseri K7(Ri15)), Fisher’s Least Significant Difference (LSD) pairwise comparison was computed using ANOVA error term. These tests have more power than common t-tests because the error term is smaller. Genes with p < 0.01 were considered as differentially expressed. Exploratory analysis on the subset of differentially expressed genes was performed using TIGR Mev software for principle component analysis (PCA) and Pearson correlation cluster analysis.

Pathway analysis was conducted using GenMAPP (11), MAPPFinder (Gene database used-Mm-Std_20070817.gdb)(12) and Metacore Version 5.4 (GeneGo, St. Joseph, USA). Additionally, GO-Elite software (25) was used to further filter and annotate the results obtained with MAPPFinder (http://www.genmapp.org/go_elite/go_elite.html). Biological pathways identified as significant had to fulfill three criteria to be reported: (1) z-score > 2; (2) p-value < 0.05; (3) at least five genes present in the pathway. The complete gene expression data will be deposited in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) according to the MIAME standard.

RESULTS

Clinical observation of the mice. No differences in physiological parameters (weight gain, feed intake, behavior, stool texture) could be observed during the duration of the trial that would highlight a clinical effect of any of the four bacterial treatments on the mice. In particular, the protocol selected for inducing an infection on the mice (strains and doses of E. coli, strain of mice, duration of treatment) did not lead to clinical symptoms of E. coli O157:H7 infection such as diarrhea.

Hematological parameters. All the hematology parameters tested (Table 1) were in accordance with reference values (ABC Vet), except for the number of platelets in the three groups that received E. coli O157:H7 where the values were slightly higher than the reference values (140-600x10^9/l) and significantly higher compared to control group (p ≤ 0.05; n = 5). The number of platelets in mice is usually higher than in other species (900-1’600x10^9/l for
mice, 200-600x10⁹/l for humans). Due to their small size their number can often be underestimated, since some hematology analyzers have the capability of detecting only larger platelets. Murine platelets have the tendency to undergo activation in in vitro conditions and subsequent clumping occurs which leads to discrepancies in the counting procedure (14). Platelet counts in our experiment may be elevated according to the reference values of the ABC Vet hematology analyzer but are in accordance with reference values from literature (100-1'000x10⁹/l) (22).

The leukocyte (WBC) values in the pre-incubation group were in accordance with the reference values, but statistically higher compared to the control group (p = 0.02; n = 5).

**TABLE 1. Hematology values for the five groups of animals in the trial**

<table>
<thead>
<tr>
<th></th>
<th>unit</th>
<th>ref. values</th>
<th>Control</th>
<th>E. coli O157:H7</th>
<th>Pre-incubation</th>
<th>Co-incubation</th>
<th>L. gasseri K7(Rif')</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBC</td>
<td>10⁹/l</td>
<td>3.0-15.0</td>
<td>5.0 ± 2.4</td>
<td>7.4 ± 2.4</td>
<td>7.6 ± 1.9</td>
<td>6.1 ± 1.7</td>
<td>6.2 ± 2.1</td>
</tr>
<tr>
<td>RBC</td>
<td>10¹²/l</td>
<td>5.0-12.0</td>
<td>9.6 ± 0.4</td>
<td>9.8 ± 4.6</td>
<td>9.6 ± 0.4</td>
<td>9.8 ± 0.5</td>
<td>9.8 ± 0.4</td>
</tr>
<tr>
<td>HGB</td>
<td>g/l</td>
<td>111-180</td>
<td>146 ± 8</td>
<td>144 ± 68</td>
<td>143 ± 5</td>
<td>147 ± 7</td>
<td>147 ± 7</td>
</tr>
<tr>
<td>HCT</td>
<td>l/l</td>
<td>0.4-0.5</td>
<td>0.5 ± 0.1</td>
<td>0.4 ± 0.2</td>
<td>0.4 ± 0.1</td>
<td>0.5 ± 0.1</td>
<td>0.5 ± 0.1</td>
</tr>
<tr>
<td>PLT</td>
<td>10⁹/l</td>
<td>140-600</td>
<td>473 ± 330</td>
<td>926 ± 268</td>
<td>863 ± 81</td>
<td>853 ± 142</td>
<td>669 ± 298</td>
</tr>
<tr>
<td>MCV</td>
<td>fl</td>
<td>44-69</td>
<td>47 ± 1</td>
<td>46 ± 23</td>
<td>47 ± 1</td>
<td>47 ± 1</td>
<td>46 ± 1</td>
</tr>
<tr>
<td>MCH</td>
<td>pg</td>
<td>12-24</td>
<td>15 ± 1</td>
<td>15 ± 7</td>
<td>15 ± 1</td>
<td>15 ± 1</td>
<td>15 ± 1</td>
</tr>
<tr>
<td>MCHC</td>
<td>g/l</td>
<td>216-420</td>
<td>322 ± 9</td>
<td>320 ± 158</td>
<td>320 ± 5</td>
<td>320 ± 7</td>
<td>325 ± 8</td>
</tr>
</tbody>
</table>

Mean ± SD of the animals (n = 5) for each of the five groups

**Manual determination of differential white blood cell count.** The values for monocytes, eosinophils and basophils each comprised less than 1% of the total number of cells (Table 2). The percentage of lymphocytes in the E. coli O157:H7 group was lower than in the control group (p < 0.05; n = 5). Also, a higher percentage of segmented and band neutrophils was detected in the E. coli O157:H7 group compared to the control group (p < 0.05; n = 5).
TABLE 2. Differential white blood count for the five groups of animals in the trial

<table>
<thead>
<tr>
<th></th>
<th>ref. values</th>
<th>Control</th>
<th>E. coli O157:H7</th>
<th>Pre-incubation</th>
<th>Co-incubation</th>
<th>L. gasseri K7(Rif')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphocytes (%)</td>
<td>90.2 ± 3.0</td>
<td>84.4 ± 8.3</td>
<td>71.4 ± 5.2</td>
<td>80.2 ± 5.0</td>
<td>83.4 ± 4.7</td>
<td>83.1 ± 4.4</td>
</tr>
<tr>
<td>Segmented neutrophils (%)</td>
<td>8.3 ± 3.8</td>
<td>14.4 ± 8.2</td>
<td>26.0 ± 5.2</td>
<td>18.9 ± 5.3</td>
<td>15.7 ± 4.7</td>
<td>16.1 ± 4.6</td>
</tr>
<tr>
<td>Band neutrophils (%)</td>
<td>n.d.</td>
<td>0.8 ± 0.4</td>
<td>2.3 ± 0.4</td>
<td>0.9 ± 0.6</td>
<td>0.8 ± 0.4</td>
<td>1.4 ± 1.1</td>
</tr>
<tr>
<td>Eosinophils (%)</td>
<td>0.7 ± 0.8</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Basophils (%)</td>
<td>0.2 ± 0.4</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Monocytes (%)</td>
<td>0.6 ± 0.8</td>
<td>0.3 ± 0.3</td>
<td>0.2 ± 0.3</td>
<td>0.1 ± 0.2</td>
<td>0.2 ± 0.3</td>
<td>0.1 ± 0.2</td>
</tr>
</tbody>
</table>

n.d.- not determined
Mean ± SD of the animals (n = 5) for each of the five groups

Phagocytosis. There were no statistically significant differences in phagocytic activity of granulocytes and monocytes from blood samples of animals in the various treatment groups. The mean values of granulocytes positive for phagocytosis were 80% (control), 81% (E. coli O157:H7), 85% (co-incubation), 86% (pre-incubation), and 81% (L. gasseri K7(Rif')). The mean values of monocytes positive for phagocytosis were 54% (control), 55% (E. coli O157:H7), 60% (co-incubation), 57% (pre-incubation), and 54% (L. gasseri K7(Rif')).

The lack of differences in phagocytic activity in blood samples of animals from the various treatment groups could be attributed to the fact that blood samples were taken one week after the last application of L. gasseri K7(Rif') and two to three weeks after application of E. coli O157:H7. Also, the selected protocol may not have been sufficiently infective to promote a measurable phagocytotic activity at this time point.

Statistical analysis of the genes differentially expressed in mice in response to E. coli O157:H7 and L. gasseri K7(Rif'). An analysis of the transcriptomes in each group revealed that, out of the 21’248 genes (Entrez GeneID) present on the microarray, 12’138 genes were detected in the blood cells of the mice in at least one of the five groups (Fig. 2). In the E. coli O157:H7 group 1’541 genes changed their expression significantly whereas, only 394 genes
changed their expression significantly in the *L. gasseri* K7(Rif') group. In these two sets of genes, 128 were differentially expressed in both groups (Fig. 2A).

Compared to the *E. coli* O157:H7 group, pre- and co-incubation of the mice inoculated with *E. coli* O157:H7 with *L. gasseri* K7(Rif') resulted in lower number of differentially expressed genes i.e. 551 genes and 1'103 genes, respectively. Out of these, 182 genes (33%) and 423 genes (38%), respectively, were also differentially expressed in the *E. coli* O157:H7 group (Fig. 2B).

*Figure 2. Number of differentially expressed genes in each of the four treated groups compared to the control group (A) Flow chart of the gene selection and number of altered genes in mice treated with *L. gasseri* K7(Rif') and *E. coli* O157:H7. (B) Venn diagram of the differentially expressed genes (p < 0.01) and the overlap of the four groups.*
A complete list of the differentially expressed genes is given as supplement information (Supplement Material 1). The 3’281 probesets that were differentially expressed in at least one of the treated groups relative to the control group (p < 0.01) were selected for PCA and Pearson correlation cluster analysis (Fig. 3). Also a dichotomous analysis was performed on proportion of non-significant probes (p < 0.05) in each treatment group.

In the PCA analysis (Fig. 3A) the first two principle components accounted for 52% of the variation in the data and allowed a clear clustering of the five groups. In addition, the following observations could be made: (a) the _E. coli_ O157:H7 group is clearly separated from the control group; (b) the _L. gasseri_ K7(Rif₄) group is close to the control group; (c) each of the three groups containing _E. coli_ O157:H7 are separated; (d) among the three groups inoculated with _L. gasseri_ K7(Rif₄) the _L. gasseri_ K7(Rif₄) group is closer to the pre-incubation group than to the co-incubation group.

Unsupervised Pearson correlation cluster analysis with the 3’281 probesets first revealed the presence of two major clusters, the _E. coli_ O157:H7 group being readily differentiated from the other four groups (Fig. 3B). These four groups are further clustered in two subgroups comprising, on one hand, the control group and the _L. gasseri_ K7(Rif₄) group and, on the other hand, the pre-incubation and co-incubation groups. Finally, the Pearson correlation cluster further separates the mice in the control and _L. gasseri_ K7(Rif₄) groups.

The dichotomous analysis was performed to assess the proportion of genes whose expression is not significantly changed in the genome of each treatment group relative to the control group (p > 0.05) (Fig. 3C). This proportion was the highest in the _L. gasseri_ K7(Rif₄) group (92%), followed by the pre-incubation group (90%), the co-incubation group (85%) and, finally, the _E. coli_ O157:H7 group (79%). Pairwise comparisons with a student t-test showed that all of these differences were statistically significant with p < 0.01.
Figure 3. Global analysis of the transcription pattern as a consequence of different treatments. A) PCA was computed on 3,281 differentially expressed probesets. Each dot in the PCA indicates one animal. The first two principle components contain 52% of variation in the data. Ellipses are drawn that circle each of the five groups, namely the E. coli O157:H7 (black dots), control (green dots), L. gasseri K7(Rif) (blue dots), pre-incubation (red dots) and co-incubation (yellow dots) groups. B) Unsupervised Pearson correlation cluster analysis of the five groups of mice. Probesets for which the expression level was above, below, and equivalent to the mean are shown in red, green, and black, respectively. The color scale gives the log2 ratio between -1 to 1. C) Dichotomous analysis on the proportion of non-significant probes (p > 0.05) in each treatment group relative to the control. A student t-test shows that the proportion of non-significant probes in the L. gasseri K7(Rif) group is significantly higher compared to the pre-incubation and co-incubation groups (p < 0.01). Similarly, the proportion of non-significant probes in the E. coli O157:H7 group is significantly lower compared to the co-incubation and pre-incubation groups (p < 0.01).
Functional analysis of the blood cell transcriptomic profiles. A functional analysis of the transcriptomic profiles identified 44 biological pathways that, compared to the control group, were significantly enriched in at least one of the four treatment groups of mice (Fig. 4). These pathways are presented groupwise in the following three sections.

Figure 4. Overview of all pathways significantly changed in at least one treatment group. The barplots show the pathways that are up-regulated or down-regulated using either GenMAPP (criteria of selection: Z score > 2; p < 0.05; number of differentially expressed genes in pathway > 5) or GeneGO network analysis (criteria for selection: -log p value > 2; number of differentially expressed genes in pathway > 5). The significant pathways are indicated by a plain red color.

Figure 5 shows the expression levels, relative to the control group, of the genes involved in the process of glycolysis/gluconeogenesis, highlighting the increased expression of this group of genes in the E. coli O157:H7 group.
Figure 5. Relative levels of expression, compared to the control group, for the genes of the glycolysis/gluconeogenesis pathway in each of the four treatment groups. Colored symbols in each group indicate genes that are significantly differentially expressed (p < 0.01). The y axis shows the log2 ratio values for the relative expression of each of the genes.

Interestingly, the TGF-β receptor signalling pathway, an important anti-inflammatory pathway in mucosal and systemic immunity, was significantly enriched in down-regulated genes (Ube2d2, Sp1, Skp1a, Rock1, Snx6, Atf2, Ets1, Foxo3a, Nup153, Ctcf, Ube2d3, Sumo1, Camk2a, Snip1, Crebbp) in the E. coli O157:H7 group (Fig. 6).
Figure 6. Relative levels of expression, compared to the control group, for the genes of the TGF-β receptor signalling pathway in each of the four treatment groups. Colored symbols in each group indicate genes that are significantly differentially expressed (p < 0.01). The y axis shows the log2 ratio values for the relative expression of each of the genes.

Taken together the modulation of these immunological pathways is in line with an inflammatory activation of the blood cells in response to the presence of *E. coli* O157:H7 in the gut and/or in the blood stream of the mice.

**Functional analysis of the differentially expressed genes in the pre- and co-incubation groups.** The second panel in Fig. 4 shows that, for almost all of the 44 biological pathways identified in the study, the number of statistically significant genes distributed into these
pathways was lower in the pre-incubation group compared to the *E. coli* O157:H7 group. Also, only 6 functional pathways (2 up-regulated, 4 down-regulated) were enriched in the pre-incubation group. In particular, the enrichment of up-regulated immunomodulatory pathways observed in the *E. coli* O157:H7 group (inflammation, immune response, cell adhesion) was not observed in the pre-incubation group.

In the co-incubation group (third panel in Fig. 4) the number of statistically significant genes distributed into the 44 pathways identified in the study was lower compared to the *E. coli* O157:H7 group but higher compared to the pre-incubation group. Furthermore, the number of statistically enriched functional pathways (10 up-regulated, 6 down-regulated) in the co-incubation group was lower compared to the *E. coli* O157:H7 group but higher compared to the pre-incubation group.

In contrast to the *E. coli* O157:H7 group that showed down-regulation of the TGF-β receptor signalling pathway, this pathway was up-regulated in the co-incubation group (see Fig. 4). In particular, a set of 9 genes belonging to this pathway was up-regulated in the co-incubation group (Dvl1, Atf2, Camk2d, Rock1, Nup153, Prkcd, Zeb1, Mapk14, Erbb2ip) (Fig. 6).

Pathways related to the complement cascade (*C1qa, C1qb, C1qc, C1r, C3arl, C4b, Cfb*) were significantly down-regulated in the co-incubation group (see Fig. 4). A closer inspection of the genes differentially expressed in these pathways reveals that a down-regulating trend was also observed in the other treatment groups, but that these effects did not reach statistical significance (Fig. 7). The complement cascade is a proteolytic system that has a paramount importance in regulating the innate immunity, in particular during the development of infectious and autoimmune diseases. Surprisingly, a role for the complement cascade in enterohemorrhagic infections as well as in the interaction of probiotic with host organisms, has not been widely reported in the literature. In that context, we note that the genes of the complement cascade show the highest fold-changes in our data set, with the log2 ratio reaching a value of almost -4 for C1qb in the co-incubation group (i.e. 16-fold decrease in gene expression).
Figure 7. Relative levels of expression, compared to the control group, for the genes of the complement cascade in each of the four treatment groups. Colored symbols in each group indicate genes that are significantly differentially expressed ($p < 0.01$). The $y$ axis shows the log2 ratio values for the relative expression of each of the genes.

**Functional analysis of the *L. gasseri* K7(Rif$^\ddagger$) group.** Oral supplementation of the mice with *L. gasseri* K7(Rif$^\ddagger$) alone also resulted in the activation of a number of functional pathways (21 up-regulated, 2 down-regulated) (fourth panel in Fig. 4). As for the *L. gasseri* K7(Rif$^\ddagger$) group, the enriched pathways belonged to immunomodulatory biological categories (inflammation, immune response, cytoskeleton, cell adhesion). However, the number of differentially expressed genes distributed in the 44 pathways identified in this study were lower in the *L. gasseri* K7(Rif$^\ddagger$) group compared to the *E. coli* O157:H7 group.

The TGF-β receptor signalling pathway was significantly up-regulated in the *L. gasseri* K7(Rif$^\ddagger$) group in contrast to the *E. coli* O157:H7 group for which down-regulation was observed (see Fig. 4). Inspection of the differentially expressed genes in the TGF-β receptor signalling pathway revealed, however, that the most striking differences in the distribution of significantly up-regulated and down-regulated genes lies between the co-incubation group,
that was also significantly enriched in up-regulated genes, and the \textit{E. coli} O157:H7 group (see Fig. 6).

\textbf{Discussion}

The pathogenic action of \textit{E. coli} O157:H7 can best be studied in antibiotic-treated (51) or germ-free (49) mice, although the pathogenicity of this strain was also demonstrated in conventional mice with an intact microflora (32, 44). One study found that conventional mice are asymptomatic for infection by \textit{E. coli} O157:H7 and concluded that this strain behaves as commensal bacteria rather than a pathogen (33). In that context, our study also did not reveal evident clinical signs of infection by \textit{E. coli} O157:H7 in C57BL/6J mice. The increased ratio of neutrophils to lymphocytes in the blood samples of the \textit{E. coli} O157:H7 group suggests a moderate inflammatory activation and WBC translocation in response to the pathogen. Despite this observation, classical clinical analyses of the blood cells (i.e. determination of hematological parameters in the context of "normal ranges") were inconclusive and a clear hematological profile of bacterial infection and/or inflammation could not be deducted. Results of phagocytic activity of granulocytes and monocytes were within normal ranges in all groups, no phagocytosis deficit or hyperactivation was therefore detected. Taken together our findings corroborate the difficulty in finding optimal experimental conditions for studying the pathogenic action of this bacterial strain in murine models.

\textit{E. coli} O157:H7 binds to the follicle-associated epithelium of Peyer’s patches in the intestine and activates the underlying human macrophages (36, 38). Similarly probiotic strains also bind to Peyer’s patches and modulate the underlying immune cells such as macrophages and dendritic cells (4, 37). These properties legitimate the research strategy in our study in which we used the blood cell transcriptome of mice to evaluate the potential of \textit{L. gasseri} K7(Rif\textsuperscript{R}) to inhibit \textit{E. coli} O157:H7 infection.

Indeed, in comparison to clinical analysis of blood, transcriptomic analysis allowed us to identify molecular changes in the blood cells of mice treated with \textit{E. coli} O157:H7 that provide quantitative and mechanistic information on the infection of these, otherwise asymptomatic, animals. Evidence for a subclinical action of \textit{E. coli} O157:H7 can indeed be derived from the analysis of several biological pathways that were activated at the gene expression level.
Firstly, *E. coli* O157:H7 adherence to host cell surface is an important determinant of virulence. *In vitro* studies have previously shown that *E. coli* O157:H7 adheres to the surface of cells such as epithelial cells (46) and macrophages (38, 47) by rearranging the host cytoskeleton. Furthermore, *E. coli* O157:H7 adherence to epithelial cells activates protein kinase C (PKC) and calmodulin (10), a signalling pathway that participates in the disruption of the intestinal barrier induced by *E. coli* O157:H7 (46). An analysis of the pathways activated in the *E. coli* O157:H7 group revealed that the focal adhesion and cytoskeleton rearrangement processes were up-regulated in the blood cells along with an increased activity of PKC, calmodulin 3, and diacylglycerol kinase.

Secondly, *E. coli* O157:H7 and effector molecules of this strain such as flagellin, initiate a strong inflammatory response in host cells (31, 50). For instance, infection of rabbits with *E. coli* O157:H7 leads to severe intestinal inflammation (24). The Shiga toxin produced by *E. coli* O157:H7 increases production of inflammatory cytokines such as TNF-alpha and IL-6 in murine macrophages (50). In a monocytic cell line, the induction of pro-inflammatory cytokines by *E. coli* O157:H7 results from the activation of signalling pathways involving MAP kinase (5). In our study, the up-regulation of JAK-STAT, and MAPK pathways in the *E. coli* O157:H7 group provides evidence for an inflammatory process.

Thirdly, the catabolism of host carbohydrates by bacterial enzymes is important for infection and pathogenesis as enteric pathogens rely on these molecules as energy sources (3, 45). Furthermore, *E. coli* O157:H7 preferentially uses glycolytic substrates as a source of carbon for its growth in mice (30). Finally, *E. coli* O157:H7 has been reported to reside inside macrophages (38). The observed up-regulation of the glycolysis/glucconeogenesis pathway in *E. coli* O157:H7 group provides further evidence for an ongoing infectious process.

Taken together the modulation of the above pathways in the group infected with *E. coli* O157:H7 strongly suggests that this pathogen has activated the blood cells of the mice. This activation may take place directly in response to a direct contact between *E. coli* O157:H7 and the blood cell in the Peyer’s patches or, following translocation of *E. coli* O157:H7 through the intestinal barrier, in the circulation. It may also take place indirectly following the activation of a systemic inflammatory process triggered by the infection on the apical side of the intestinal wall. That the glycolysis/glucconeogenesis pathway is significantly up-regulated in the *E. coli* O157:H7 group provides non-exclusive evidence for the first hypothesis.
Gene expression profiling of blood cells clearly allows us to differentiate mice inoculated with *E. coli* O157:H7 from mice inoculated with *L. gasseri* K7(Rif'). For example, 13% of the genes on the arrays were differentially expressed in the *E. coli* O157:H7 group whereas, only 3% of the genes were modulated by *L. gasseri* K7(Rif'). These two numbers can be brought in the context of the much higher number of *L. gasseri* K7(Rif') cells that were inoculated in the mice (14 daily doses of $1 \times 10^9$ cells) compared to *E. coli* O157:H7 (one single dose of $5 \times 10^8$ cells). PCA and Pearson correlation analyses not only differentiate both treatment groups, but further suggest that the *L. gasseri* K7(Rif') group is biologically closer to the control group than the *E. coli* O157:H7 group. Finally, an analysis of the enriched biological pathways in both groups also provides evidence that the blood cells of mice inoculated with *L. gasseri* K7(Rif') are transcriptionally less active, compared to the control group, than the mice inoculated with *E. coli* O157:H7. Taken together these two observations suggest that the blood cell transcriptome is very sensitive in picking up subtle but biologically relevant differences in gene expression which could differentiate the properties of *E. coli* O157:H7 and *L. gasseri* K7(Rif') in their ability to interact with the host. In particular, the blood cell transcriptome may conceptually differentiate between a pathogen–host response (19) and a commensal-host response (23).

Although the mice infected with *E. coli* O157:H7 were clinically asymptomatic, several observations on the blood cell transcriptome suggest that *L. gasseri* K7(Rif') was able to modify the action of *E. coli* O157:H7. Firstly, both co- and pre-incubation of the mice infected with *E. coli* O157:H7 with *L. gasseri* K7(Rif') lowered the number of differentially expressed genes. Consequently, the number of statistically-enriched pathways was also significantly reduced. Secondly, PCA and Pearson correlation clustering analyses were clearly able to differentiate the co- and pre-incubation groups from the *E. coli* O157:H7 group. Taken together the PCA, Pearson correlation clustering and dichotomous analyses show that oral administration of *L. gasseri* K7(Rif') statistically moved the blood cell transcriptome of the mice infected with *E. coli* O157:H7 closer to the control group and that pre-incubation was more efficient than co-incubation in doing so.

The increased ratio of neutrophils to lymphocytes observed in the *E. coli* O157:H7 group suggests an activation of innate immunity and may thus participate in the transcriptomic profile observed in this group of mice, for example by reflecting an activation of the antibacterial properties of neutrophils. This observation grants further studies in which the
gene expression profiles of isolated populations of blood cells, in particular neutrophils, would be measured. Meanwhile, our transcriptomic analysis clearly goes beyond the mere counting of immune cells by performing a clustering analysis of the five treatments groups based on the differentially expressed genes and by providing mechanistic information on the pathways modulated by these genes in the blood cells of the inoculated mice.

In that context, the profiles of the biological pathways identified in the different groups of mice in Fig. 4 are dominated by the pathways enriched in the E. coli O157:H7 group. Consequently, the significant differences detected between the various groups are more quantitative than qualitative in nature and few, if any, pathways are uniquely enriched, compared to the E. coli O157:H7 group, in the three groups inoculated with L. gasseri K7(Rif)]. In that context, it is worth noticing that the TGF-β receptor signalling is down-regulated in the E. coli O157:H7 group but up-regulated in the L. gasseri K7(Rif) group and in the co-incubation group. TGF-β receptor signalling is important in establishing intestinal homeostasis and prevention against pathogens, such as E. coli O157:H7, by promoting the synthesis of IgA by immune cells (2, 9) and by activating anti-inflammatory signalling (7, 52). E. coli O157:H7 disrupts the integrity of the epithelial barrier, a property that is inhibited by the presence of TGF-β in cultures of epithelial cells (20). Also, elevated levels of circulating TGF-β are found in children infected with E. coli O157:H7 but with reduced incidence of hemolytic uremic syndrome (39). In our study, the over-representation in the E. coli O157:H7 group of the down-regulated genes of the TGF-β receptor signalling pathway was reduced upon co-incubation with L. gasseri K7(Rif) and might indicate an activation of the IgA response against E. coli O157:H7 in this group of animals. Immuno-histological analyses of intestinal sections on these mice support this conclusion (Hacin et al., manuscript in preparation).

In summary, our transcriptomic study on mice challenged with E. coli O157:H7 shows that L. gasseri K7(Rif) inhibits the blood cell transcriptomic profile induced by E. coli O157:H7, a profile that is mainly characterized by an activation of immune-modulatory pathways. Our data also suggest that a pre-inoculation protocol, in which L. gasseri K7(Rif) is administrated one week prior to the infection by E. coli O157:H7 is more effective than a co-incubation protocol. On the technical side, our study provides evidence that blood cells transcriptomic profiling can be used as a molecular diagnostic and exploratory tool to study bacterial infections in vivo under sub-clinical asymptomatic conditions.
ACKNOWLEDGEMENT

This work was supported by Agroscope Liebefeld-Posieux research station ALP internal funding and by the Slovenian Research Agency (Research program P4-0097).

We thank Dr. B. Walther, Dr. R. Sieber, A. Schmid, and D. Gille (Berne, Switzerland) for critical comments on the manuscript; C. Aquino (FGCZ, Zurich, Switzerland) for services on the microarray experiment, Dr. W. Luginbühl (ChemStat, Berne, Switzerland) for statistical analysis.

F. Sagaya (design, conductance, and interpretation of the microarray experiment; redaction of the manuscript) and B. Hacin (design and conductance of the experimental part of the clinical mouse study) equally contributed to this work and should both be acknowledged as first authors. G. Tompa, A. Ilhan, S. Škrajnar, M. Černe, B. B. Matijašić, I. Rogelj were involved in designing or conducting the experimental part of the clinical mouse study. F. Sagaya, R. F. Hurrell and G. Vergères were involved in design of the microarray experiment and in writing the manuscript. None of the authors has a personal or financial conflict of interest.
Reference


CONCLUSION

Conclusion and future perspective

Nutrigenomics is a relatively new branch of nutrition and food research that is targeted towards a better understanding of the interaction between the genome and its nutritional environment by applying new high throughput profiling techniques, in particular the ‘omics’ technologies. One of the main advantage of these technologies, such as microarrays, is the ability to obtain a holistic view of molecular events taking place in living organisms. However, despite the use of attractive technologies nutrigenomics studies face various challenges in collecting data, analyzing them, and deducing meaningful biological interpretations (Wittwer et al., 2011). One such challenge in data collection is the accessibility of tissues responsive to dietary stimuli by noninvasive methods. In this thesis, blood cell transcriptomics was applied (a) in a human nutritional intervention study to investigate the postprandial effect of milk and yogurt ingestion and (b) to investigate the antimicrobial effect of a potential probiotic strain of Lactobacillus gasseri on E. coli infection in a mouse model.

In the human nutritional intervention study, two major biological effects, namely a metabolic effect and an immunomodulatory effect, were identified in blood cells after the ingestion of milk and yogurt by healthy volunteers. At the metabolic level, genes involved in protein synthesis and mitochondrial functions, such as ribosomal genes and genes involved in oxidative phosphorylation, were up-regulated. In addition several anti-oxidant genes, such as SOD1, TXN2 and selenoproteins, were up-regulated. At the immunomodulatory level, genes involved in the inflammatory response, such as TLR4, TLR2, NF-κB, IL6 and IL1b, were down-regulated. Furthermore, a transcription factor enrichment analysis revealed that binding sites for NRF2, which activates the expression of genes involved in the anti-oxidative response, were enriched in the group of up-regulated genes. Overall, our study indicates that, in human adults, the ingestion of bovine milk and yogurt may mitigate the postprandial inflammatory response by increasing the anti-oxidative response.

Most of the regulatory properties measured after the ingestion of milk were also observed with yogurt, in agreement with the shared macronutrient composition of the two dairy products. Interestingly, in spite of these overall similarities, several differences were observed
between milk and yogurt. One of the striking differences was a significant down-regulation of the ACE2 pathway in the yogurt group, a pathway that has implication in reducing blood pressure and that is already discussed for fermented dairy products in the literature.

One of the novel aspect of this work was cross-over design with a kinetic analysis of gene expression changes over a postprandial time window of six hours. This experimental design not only minimized inter-individual variation but also allowed the identification of two groups of genes with different biological function and specific kinetic patterns of gene expression (e.g. immunomodulatory genes with up-DOWN kinetics and genes involved in mitochondrial function or protein synthesis with down-UP kinetics). These kinetic patterns could thus serve as class discovery models and lead to the identification of biomarkers.

Future work building on the data from this class comparison study could focus along the following lines: (i) validating the observations with a large and mixed population, (ii) investigating long term effects of milk and yogurt, and (iii) testing the main hypothesis generated by this work that milk and yogurt have an anti-inflammatory property at the protein and cellular level. Indeed, follow-up research investigating the last two points in conditions of increased inflammation and oxidative stress i.e. in overweight and obese human subjects, is currently ongoing in the context of the NutriChip project. Readers may refer to the following site for more information on this project: http://www.nanoterachoprojects/403.php.

The mouse study revealed that blood cell transcriptomic profiling is able to detect molecular events leading to bacterial infections under sub-clinical asymptomatic conditions. A descriptive analysis of the entire gene expression data set showed that L. gasseri K7(Rif⁶) induced a gene expression profile in blood cells of mice that is similar/closer to the control group of mice and that is distinctively different from the profile induced by E. coli O157:H7. A functional analysis of the differentially expressed genes in mice treated with E. coli O157:H7 showed modulation of pathways that are typical of bacterial infection, more specifically cell adhesion, cytoskeleton rearrangement, inflammation, and glycolysis. While, L. gasseri K7(Rif⁶) group showed an expression profile that is close to the control group of untreated mice, the gene expression profile induced by E. coli O157:H7 was inhibited by pre-incubating or co-incubating these mice with L. gasseri K7(Rif⁶). In addition, the pre-incubation protocol seemed to be more effective than the co-incubation protocol in limiting the infective response induced by E. coli O157:H7.
In conclusion, our studies demonstrated that blood cell transcriptomics is a sensitive and effective tool for an holistic evaluation of the molecular in vivo response of mammalians to dairy products, including lactic acid bacteria.
CURRICULUM VITAE

Sagaya Mary Francina
born 16.10.1975
in India

Education

2006-2011
PhD Thesis in the field of Nutrigenomics
Laboratory of Human Nutrition
Institute of food science and technology

2002
Postgraduate Diploma in Bioinformatics
SISI Guindy, Chennai, India

1996-1998
Masters in Molecular Biology
University of Madras, India

1993-1996
Bachelors in Zoology
Stella Maris College,
University of Madras, India

1991-1993
Higher secondary school
Central schools (Kendriya Vidyalaya), India

1980-1991
Primary and secondary school
Central schools (Kendriya Vidyalaya), India

Professional experience

2002-2004
Research Scientist
Gemeiosys Research Labs India Pvt. Ltd.
Chennai, India