Integrative Analysis of the Genes Induced by the Intestine Microbiota of Infant Born to Term and Breastfed

Badreddine Nouadi, Yousra Sbaoui, Mariame El Messal, Faiza Bennis and Fatima Chegdani

Laboratory of Health and Environment, Faculty of Sciences Ain Chock, Hassan II University of Casablanca, Casablanca, Morocco.

ABSTRACT: Nowadays, the integration of biological data is a major challenge for bioinformatics. Many studies have examined gene expression in the epithelial tissue in the intestines of infants born to term and breastfed, generating a large amount of data. The integration of these data is important to understand the biological processes involved during bacterial colonization of the newborns intestine, particularly through breast milk. This work aims to exploit the bioinformatics approaches, to provide a new representation and interpretation of the interactions between differentially expressed genes in the host intestine induced by the microbiota.

KEYWORDS: Gene expression, network, intestinal microbiota, newborn, breastfed

A total of 61 differentially expressed genes (DEGs) in the intestine of newborns extracted from several bibliographic works and databases were annotated for functional analysis using the String software (http://string-db.org/), the Cytoscape software (http://www.cytoscape.org/) and the BiNGO plugin. The latter provided an evaluation of the signaling and metabolic pathways, molecular networks and biological processes for all the used genes.

The analysis revealed that RELA, INS, IRS1, IL1B, and NFKBIA are the central genes in the interaction networks produced. These networks show that the cellular differentiation of the intestinal epithelium and the development of mucosal immunity, are the most affected processes in newborns. Therefore, the global patterns of interactions supports the relationship between breastmilk and the role of the microbiota’s diversity. Ergo all results consolidate the importance of breastmilk and intestinal microbiota in homeostasis.

Introduction
Breastmilk microbiota provides the transient microbiota in the newborn. This transient microbiota is important for the implantation of the personalized intestinal microbiota of each individual; indeed, it plays a fundamental role in the development of the newborn. Moreover, breastfeeding has demonstrated its ability to provide a balanced intestinal microbiota to the newborn, thus positively impacting the newborn’s health. Human milk can stimulate the proliferation of bifidobacterium and lactobacillus strains, whose role is to create an acidic environmental rich in short-chain fatty acids (SCFAs) with a protective and nutritive role at the intestinal level. It has been shown also, that in germ-free rats, Streptococcus thermophilus (transient commensal bacteria) induces the epithelial stem cell differentiation. As well, Bacteroides, very abundant in human colostrum, may have a main role in the early stages of newborn’s gut colonization.

Intestinal microbiota is largely studied since 2 decades, and a big amount of data are generated with Omics approaches. However, the interaction between human milk microbiota and newborn’s intestinal microbiota is not completely clear. Other studies are needed to understand the powerful relationship between the human milk microbiota and the stimulation of newborn’s homeostasis. Novel approaches have been developed to study the microbiota using the directed acyclic graphs (DAG) method to facilitate an understanding of the ontological profiles provided by the interaction between the induced genes in the newborn’s epithelium breastfeeding.

Ontological studies are becoming essential to understand the complex mechanism placed in the intestine of newborn during breastfeeding. Thus, the biological processes and Top functions can be identified through the predicted networks.

Materials and Methods
This work consists of studying, through a statistical-computing approach, the representation of biological data to make their integration more efficient in the case of DEGs in the newborn's intestine. For this purpose, we have chosen the scientific
and 16 genes remain outside this network. A set of 51 proteins interaction between the proteins (Figure 2). Among the 67 evidence indexes (Figure 1) and a network with the different other 10 genes enriched the networks generated. The results by the String software. Fifty-seven genes were annotated and String results

Results and Discussion

plugin function.

process” was chosen as a query for the analysis with the BiNGO profile as DAG interactions. The ontological level “biological zation. The network personalization determines a functional were analyzed by the BiNGO function for network personali- tomized according to the values of the chosen parameters:—A

Coexpression analysis by String software

This analysis was performed by 2 functions:

1. Visualization of interactions between genes by empha-
sizing particular criterias such as co-occurrence, coex-
pression, experimental evidence, existing databases, and

text mining.

2. Rich statistical analysis indicates that the terms are clas-
sified by their enriched P value. The P value is calculated
by a hypergeometric test and then corrected for multiple
tests using the Benjamini and Hochberg method.

Ontological analysis by Cytoscape software

The networks generated by string were imported as a pre-
existing unformatted array in Cytoscape software.

The network analyzer plugin function provides a network customization. The size and color of the nodes have been cus-

tomized according to the values of the chosen parameters:—A

The genes annotated by Cytoscape and selected manually were analyzed by the BiNGO function for network personali-

zation. The network personalization determines a functional profile as DAG interactions. The ontological level “biological process” was chosen as a query for the analysis with the BiNGO plugin function.

Results and Discussion

String results

Sixty-one DEGs (Tables 1 to 6) were imported and analyzed by the String software. Fifty-seven genes were annotated and other 10 genes enriched the networks generated. The results were performed in 2 formats: a network with different confi-
dence indexes (Figure 1) and a network with the different interactions between the proteins (Figure 2). Among the 67 genes annotated on String, 51 have formed a single network and 16 genes remain outside this network. A set of 51 proteins was found to be linked either directly or indirectly through one or more interacting proteins, suggesting the existence of func-
tional links between them (Figures 1 and 2).

The central proteins of this network (Figures 1 and 2) are as follows: IL1β, RELA, INS, IRS1, and NFKBIA.

IL-1 (interleukin-1) production (Figure 1) is mainly regu-

lated by the inflammasome, a multimeric protein complex assembled in response to various inflammatory triggers such as danger signals, microbial toxins, and crystalline substances.\textsuperscript{20-22} A prototypical complex of inflammasome includes many pro-
teins, among them CASP1 (caspase-1). Cleavage of CASP1 by the inflammasome leads to its activation, which in turn cleaves IL-1β (IL-1 beta).\textsuperscript{23,24} Interleukin-1β, a proinflamatory cytokine with a wide range of systemic and local effects\textsuperscript{25} can modulate the function of both immune and nonimmune cells. Interleukin-1β also promotes T-cell activation and survival\textsuperscript{26} and works with other proinflammatory cytokines such as IL-33 (Figure 1) to promote epithelial restoration, repair, and mucosal healing in the intestine.\textsuperscript{27} Interleukin-1β can also induce the positive regulation of RELA (transcription factor p65) (Figure 1) and subsequently the activation of the canonical NF-κB pathway, which is necessary for homeostatic regulation of cell death and division in intestinal epithelia, as well as for protec-
tion against development of severe acute inflammation of intestines.\textsuperscript{28-30} Insulin receptor (INSR) mediates the pleio-
tropic actions of INS (insulin) (Figure 1). Insulin-binding leads to phosphorylation of several intracellular substrates, including IRS1 (insulin receptor substrate 1), subsequently inducing various bioactivities such as growth, differentiation, survival, increased anabolism, and decreased catabolism in many types of cells.\textsuperscript{31-33}

On the contrary, SOCS3 (suppressor of cytokine signaling 3) is regulated by several proteins within the network (Figure 1). It has an impact on multiple signaling pathways, and it is a mediator key of mucosal homeostasis. Suppressor of cytokine signaling 3 is a tumor suppressor, limiting the proliferation of intestinal epithelial cells (angiotensin-converting enzyme [ACE]) in cases of acute inflammation and tumor growth, and plays a role in wound repair.\textsuperscript{34,35} A group of reactive proteins (black lines) is formed of: ATP5B, ATP5A1, ATP5D, ATP5H, ATP5C1, ATP5G3, and ATP5O, (Figure 1) which have a common role in displaying energy and cells communications.\textsuperscript{36} The interconnected proteins within this network (Figure 1) reveal the antiproliferative effect of breastfeeding on the cells of the intestinal epithelium of newborns and the positive effect on cell differentiation. These observations also suggest the involvement of these proteins in metabolism, cell survival, and mucosal homeostasis.

The biological process analysis (Table 7) revealed the presence of different processes, significantly implicated in this net-

work (P value <0.05). The most important processes were:—positive regulation of NF-kappaB transcription factor activity (GO: 0051092);—positive regulation of cellular process (GO: 0048522);—positive regulation of lipid metabolic

Bioinformaticas and Biology Insights
| ID  | Abbreviations | Genes  | Description |
|-----|---------------|--------|-------------|
| 337 | ApoA4         | Apolipoprotein A4 | Secreted by the small intestine on chylomicrons in the intestinal lymph in response to the absorption of fats. Numerous physiological functions have been attributed to ApoA4, including a role in chylomicron assembly and lipid transport. Differentially expressed genes (DEGs) in the gut of neonates born to term and breastfed, selected from various databases and scientific publications. | |
| 335 | ApoA1         | Apolipoprotein A1 | This gene encodes apolipoprotein A1, which is the major protein component of high-density lipoprotein (HDL) in plasma. Mutations in this gene are associated with HDL deficiencies, including Tangier disease, and with systemic non-atherosclerotic cardiovascular disease. Numerous physiological functions have been attributed to ApoA1. | |
| 427 | ASAH1         | N-Acylsphingosine amide hydrolase 1 | This gene encodes a member of the acid ceramidase family of proteins. This lysosomal enzyme, which catalyzes the degradation of ceramide into sphingosine and fatty acids, has been implicated in a variety of cellular processes, including inflammatory responses and cell survival. |
| 29856 | CERS2 | Ceramide synthase 2 | This gene encodes a protein that has sequence similarity to yeast ceramide synthase. CERS2 is involved in ceramide biosynthesis and has been implicated in the regulation of cell growth and lipid metabolism. |
| 4534 | MTM1 | Myotubularin 1 | This gene encodes a dual-specificity phosphatase that acts on both phosphotyrosine and phosphothreonine. MTM1 has been implicated in muscle cell differentiation and mutations in this gene have been identified in X-linked myotubular myopathy. |
| 123 | PLEKH2 | Perilipin 2 | The protein encoded by this gene belongs to the perilipin family, members of which coat triacylglycerol droplets in lipid storage droplets. PLEKH2 has been implicated in the regulation of lipid metabolism in various tissues. |
| 834 | CASP1 | Caspase 1 | This gene encodes a protein which is a member of the caspase family of cysteinyl aspartate proteases. Caspase-1 plays a central role in the execution phase of cell apoptosis. |
| 80341 | BPIF2 | BPI fold containing family 5, member 2 | This gene encodes a protein which is a member of the BPI fold containing family of proteins. BPIF2 has been implicated in the regulation of innate immunity and inflammation. |
| 240 | ALOX5 | Arachidonate 5-lipoxygenase | This gene encodes a protein which is a member of the lipoxygenase family of enzymes. ALOX5 is involved in the synthesis of leukotrienes from arachidonic acid, which are important mediators of inflammation and allergic reactions. |
| 7410 | VAV2 | Vav guanine nucleotide exchange factor 2 | This gene encodes a protein which is a member of the Vav guanine nucleotide exchange factor family. VAV2 plays a role in the regulation of lymphocyte activation and proliferation. |
| 6668 | SP2 | Vav guanine nucleotide exchange factor 2 | This gene encodes a protein which is a member of the Vav guanine nucleotide exchange factor family. SP2 plays a role in the regulation of lymphocyte activation and proliferation. |

**Table 1.** Differentially expressed genes (DEGs) in the gut of neonates born to term and breastfed, selected from various databases and scientific publications.
| ID   | ABBREVIATIONS | GENES                  | DESCRIPTION                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------|---------------|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3553 | IL1B          | Interleukin 1 beta     | The protein encoded by this gene is a member of the interleukin 1 cytokine family. This cytokine is produced by activated macrophages as a proprotein, which is proteolytically processed to its active form by caspase 1 (CASP1/ICE). This cytokine is an important mediator of the inflammatory response and is involved in a variety of cellular activities, including cell proliferation, differentiation, and apoptosis [1](https://www.ncbi.nlm.nih.gov/gene/3553) |
| 3383 | ICAM1         | Intercellular adhesion molecule 1 | This gene encodes a cell surface glycoprotein which is typically expressed on endothelial cells and cells of the immune system. It binds to integrins of type CD11a/CD18 or CD11b/CD18 [1](https://www.ncbi.nlm.nih.gov/gene/3383)                                                                                                                                                                                                 |
| 90865| IL33          | Interleukin 33         | The protein encoded by this gene is a cytokine that binds to the IL1RL1/ST2 receptor. The encoded protein is involved in the maturation of Th2 cells and the activation of mast cells, basophils, eosinophils, and natural killer cells [1](https://www.ncbi.nlm.nih.gov/gene/90865)                                                                                                      |
| 50506| DUOX2         | Dual oxidase 2         | The protein encoded by this gene is a glycoprotein and a member of the NADPH oxidase family. The synthesis of thyroid hormone is catalyzed by a protein complex located at the apical membrane of thyroid follicular cells. This complex contains an iodide transporter, thyroperoxidase, and a peroxide-generating system that includes this encoded protein and DUOX1 [1](https://www.ncbi.nlm.nih.gov/gene/50506) |
| 51348| KLRF1         | Killer cell lectin like receptor F1 | KLRF1, an activating homodimeric C-type lectin-like receptor (CTLR), is expressed on nearly all natural killer (NK) cells and stimulates their cytotoxicity and cytokine release [1](https://www.ncbi.nlm.nih.gov/gene/51348)                                                                                                                                                         |
| 4790 | NF-kB         | Nuclear factor kappa B subunit 1 | Nuclear factor kappa B (NFkB) is a transcription regulator that is activated by various intracellular and extracellular stimuli such as cytokines, oxidant-free radicals, ultraviolet irradiation, and bacterial or viral products. Activated NFkB translocates into the nucleus and stimulates the expression of genes involved in a wide variety of biological functions. Inappropriate activation of NFkB has been associated with a number of inflammatory diseases while persistent inhibition of NFkB leads to inappropriate immune cell development or delayed cell growth [1](https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=4790) |
| 5468 | PPAR γ       | Peroxisome-proliferator-activated receptor gamma | The protein encoded by this gene is PPAR-gamma and is a regulator of adipocyte differentiation [1](https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=5468)                                                                                                                                                                                                       |
| 5465 | PPAR α       | Peroxisome-proliferator-activated receptor alpha | Peroxisome-proliferator-activated receptors (PPARs) affect the expression of target genes involved in cell proliferation, cell differentiation, and in immune and inflammation responses. This gene encodes the subtype PPAR-alpha, which is a nuclear transcription factor [1](https://www.ncbi.nlm.nih.gov/gene/5465)                                                                                                                                 |
| 4869 | NPM1          | Nucleophosmin 1       | The protein encoded by this gene is involved in several cellular processes, including centrosome duplication, protein chaperoning, and cell proliferation. The encoded phosphoprotein shuttles between the nucleolus, nucleus, and cytoplasm, chaperoning ribosomal proteins and core histones from the nucleus to the cytoplasm. This protein is also known to sequester the tumor suppressor ARF in the nucleolus, protecting it from degradation until it is needed [1](https://www.ncbi.nlm.nih.gov/gene/4869) |
| 3156 | HMGCR         | 3-Hydroxy-3-methylglutaryl-CoA reductase | 3-Hydroxy-3-methylglutaryl (HMG)-CoA reductase is the rate-limiting enzyme for cholesterol synthesis and is regulated via a negative feedback mechanism mediated by sterols and nonsterol metabolites derived from mevalonate, the product of the reaction catalyzed by reductase. Normally in mammalian cells this enzyme is suppressed by cholesterol derived from the internalization and degradation of low density lipoprotein (LDL) via the LDL receptor [1](https://www.ncbi.nlm.nih.gov/gene/3156) |

Abbreviation: ARF, alternative reading frame.
Table 3. Differentially expressed genes (DEGs) in the gut of neonates born to term and breastfed, selected from various databases and scientific publications (continued).

| ID  | ABBREVIATIONS | GENES                                      | DESCRIPTION                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----|----------------|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3936| LCP1           | Lymphocyte cytosolic protein 1             | Plasmins are a family of actin-binding proteins that are conserved throughout eukaryote evolution and expressed in most tissues of higher eukaryotes. In humans, 2 ubiquitous plasmin isoforms (L and T) have been identified. Plasmin 1 (otherwise known as fimbrin) is a third distinct plasmin isoform which is specifically expressed at high levels in the small intestine (https://www.ncbi.nlm.nih.gov/gene/3936) |
| 4792| NFKBIA         | NFKB inhibitor alpha                       | This gene encodes a member of the NF-kappa-B inhibitor family, which contain multiple ankrin repeat domains. The encoded protein interacts with REL dimers to inhibit NF-kappa-B/REL complexes which are involved in inflammatory responses (https://www.ncbi.nlm.nih.gov/gene/4792)                                                                                                                         |
| 5209| PFKFB3         | 6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 | The protein encoded by this gene belongs to a family of bifunctional proteins that are involved in both the synthesis and degradation of fructose-2,6-bisphosphate, a regulatory molecule that controls glycolysis in eukaryotes. This protein is required for cell cycle progression and prevention of apoptosis. It functions as a regulator of cyclin-dependent kinase 1, linking glucose metabolism to cell proliferation and survival in tumor cells (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=5209) |
| 6280| S100-A9        | S100 calcium-binding protein A9           | The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=6280) |
| 9021| SOCS3          | Suppressor of cytokine signaling 3         | This gene encodes a member of the STAT-induced STAT inhibitor (SSI), also known as suppressor of cytokine signaling (SOCS), family. SSI family members are cytokine-inducible negative regulators of cytokine signaling. The expression of this gene is induced by various cytokines, including IL6, IL10, and interferon (IFN)-gamma. The protein encoded by this gene can bind to JAK2 kinase, and inhibit the activity of JAK2 kinase (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=9021) |
| 54210| TREM-1        | Triggering receptor expressed on myeloid cells 1 | This gene encodes a receptor belonging to the Ig superfamily that is expressed on myeloid cells. This protein amplifies neutrophil and monocyte-mediated inflammatory responses triggered by bacterial and fungal infections by stimulating release of proinflammatory chemokines and cytokines, as well as increased surface expression of cell-activation markers (https://www.ncbi.nlm.nih.gov/gene/54210) |
| 7305| TYROBP         | TYRO protein tyrosine kinase-binding protein | This gene encodes a transmembrane signaling polypeptide which contains an immunoreceptor tyrosine-based activation motif (ITAM) in its cytoplasmic domain. The encoded protein may associate with the killer-cell inhibitory receptor (KIR) family of membrane glycoproteins and may act as an activating signal-transduction element (https://www.ncbi.nlm.nih.gov/gene/7305) |
| 1051| CEBPB          | CCAAT enhancer binding protein beta        | This intronless gene encodes a transcription factor that contains a basic leucine zipper (bZIP) domain. Activity of this protein is important in the regulation of genes involved in immune and inflammatory responses, among other processes (https://www.ncbi.nlm.nih.gov/gene/1051) |
| 8837| CFLAR           | CASP8 and FADD like apoptosis regulator    | The protein encoded by this gene is a regulator of apoptosis and is structurally similar to caspase-8. However, the encoded protein lacks caspase activity and appears to be itself cleaved into 2 peptides by caspase-8 (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=8837) |
| 9750| FAM65B         | RHO family interacting cell polarization regulator 2 | This gene encodes an atypical inhibitor of the small G protein RhoA. Inhibition of RhoA activity by the encoded protein mediates myoblast fusion and polarization of T-cells and neutrophils. The encoded protein is a component of hair cell stereocilia that is essential for hearing. A splice site mutation in this gene results in hearing loss in human patients (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=9750) |
| ID    | ABBREVIATIONS | GENES                | DESCRIPTION                                                                                                                                                                                                 |
|-------|----------------|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5265  | SERPINA1       | Serpin family A member 1 | The protein encoded by this gene is secreted and is a serine protease inhibitor whose targets include elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator [https://www.ncbi.nlm.nih.gov/gene/5265] |
| 57126 | CD177          | CD177 molecule       | This gene encodes a glycosyl-phosphatidylinositol (GPI)-linked cell surface glycoprotein that plays a role in neutrophil activation. The protein can bind platelet endothelial cell adhesion molecule-1 and function in neutrophil transmigration [https://www.ncbi.nlm.nih.gov/gene/57126] |
| 3429  | IFI27          | Interferon inducible protein 27 | Among its related pathways are cytokine signaling in immune system and innate immune system. Gene ontology (GO) annotations related to this gene include RNA polymerase II activating transcription factor binding and lamin binding [http://www.genecards.org/cgi-bin/carddisp.pl?gene=IFI27] |
| 10379 | IRF9           | Interferon regulatory factor 9 | Among its related pathways are immune response IFN-gamma-signaling pathway and cytokine signaling in immune system. Gene ontology (GO) annotations related to this gene include DNA-binding transcription factor activity [http://www.genecards.org/cgi-bin/carddisp.pl?gene=IRF9] |
| 3937  | LCP2           | Lymphocyte cytotoxic protein 2 | This gene encodes an adapter protein that acts as a substrate of the T-cell antigen receptor (TCR)-activated protein tyrosine kinase pathway. The encoded protein associates with growth factor receptor bound protein 2 and is thought to play a role in TCR-mediated intracellular signal transduction. A similar protein in mouse plays a role in normal T-cell development and activation [https://www.ncbi.nlm.nih.gov/gene/3937] |
| 10057 | ABCC5          | ATP-binding cassette subfamily C member 5 | The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extracellular and intracellular membranes. ABC genes are divided into 7 distinct subfamilies (ABC1, MDR/MDR, MRP, ALD, OABP, GCN20, and white). This protein is a member of the MRP subfamily which is involved in multidrug resistance. This protein functions in the cellular export of its substrate, cyclic nucleotides [https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=10057] |
| 506   | ATP5B          | ATP synthase F1 subunit beta | This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation [https://www.ncbi.nlm.nih.gov/gene/506] |
| 663   | BNIP2          | BCL2-interacting protein 2 | This gene is a member of the BCL2 adenovirus E1B 19 kDa-interacting protein (BNIP) family. It interacts with the E1B 19 kDa protein, which protects cells from virally induced cell death. The encoded protein also interacts with E1B 19 kDa-like sequences of BCL2, another apoptotic protector [https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=663] |
| 1030  | CDKN2B         | Cyclin-dependent kinase inhibitor 2B | This gene lies adjacent to the tumor suppressor gene CDKN2A in a region that is frequently mutated and deleted in a wide variety of tumors. This gene encodes a cyclin-dependent kinase inhibitor, which forms a complex with CDK4 or CDK6 and prevents the activation of the CDK kinases, thus the encoded protein functions as a cell growth regulator that controls cell cycle G1 progression. The expression of this gene was found to be dramatically induced by transforming growth factor (TGF) beta, which suggested its role in the TGF-beta-induced growth inhibition [https://www.ncbi.nlm.nih.gov/gene/1030] |
| 8655  | DYNLL1         | Dynein light chain LC8-type 1 | The complex is involved in intracellular transport and motility. The protein described in this record is a light chain and exists as part of this complex but also physically interacts with and inhibits the activity of neuronal nitric oxide synthase. Binding of this protein destabilizes the neuronal nitric oxide synthase dimer, a conformation necessary for activity, and it may regulate numerous biologic processes through its effects on nitric oxide synthase activity [https://www.ncbi.nlm.nih.gov/gene/8655] |
| ID      | ABBREVIATIONS | GENES                      | DESCRIPTION                                                                                                                                                                                                                     |
|---------|---------------|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 100463482 | MT-RNR2-L6   | MT-RNR2-Like 6            | It is unclear if this is a transcribed protein-coding gene, or if it is a nuclear pseudogene of the mitochondrial MT-RNR2 gene (https://www.ncbi.nlm.nih.gov/gene/100463482)                                                      |
| 5170    | PDPK1         | 3-Phosphoinositide        | Among its related pathways are constitutively signaling by AKT1 E17K in cancer and NFAT, and protein tyrosine kinase activity. Among its related pathways are ADP signaling and protein transport across epithelia in many organs. The gene encodes a protein that is associated with pseudohypoaldosteronism type 1 (pHA1), a rare salt-wasting disease resulting from target organ unresponsiveness to mineralocorticoids (http://www.genecards.org/cgi-bin/carddisp.pl?gene=SCNN1A) |
| 5911    | RAP2A         | RAP2A, member of RAS oncogene family | Among its related pathways are NFAT and cardiac hypertrophy. Gene ontology (GO) annotations related to this gene include GTP binding and GTPase activity. An important paralog of this gene is RAP2C (http://www.genecards.org/cgi-binocarddisp.pl?gene=RAP2C) |
| 6337    | SCNN1A        | Sodium channel epithelial 1 alpha subunit | Among its related pathways are ADP signaling and protein transport across epithelia in many organs. The gene encodes a protein that is associated with pseudohypoaldosteronism type 1 (pHA1), a rare salt-wasting disease resulting from target organ unresponsiveness to mineralocorticoids (http://www.genecards.org/cgi-bin/carddisp.pl?gene=SCNN1A) |
| 6670    | Sp3           | Sp3 transcription factor   | This gene belongs to a family of Sp1-related genes that encode transcription factors that either stimulate or repress the transcription of numerous genes (https://www.ncbi.nlm.nih.gov/gene/6670) |
| 55521   | TRIM36        | Tripartite motif containing 36 | This gene is a protein-containing DNA-binding and architectural factor that regulates transcription factors (http://www.genecards.org/cgi-binocarddisp.pl?gene=TRIM36) |
| 3552    | IL1A          | Interleukin 1 alpha        | The protein encoded by this gene is a member of the interleukin 1 cytokine family. This protein is an inflammatory cytokine that induces apoptosis, and plays a role in the survival of T cells (https://www.ncbi.nlm.nih.gov/gene/3552) |
| 6869    | TACR1         | Tachykinin receptor 1      | This gene belongs to a family of tachykinin receptors. The tachykinin receptors are characterized by their response to cell injury (https://www.ncbi.nlm.nih.gov/gene/6869) |
| 5966    | REL           | REL proto-oncogene, NF-κB subunit | This gene encodes a protein that belongs to the Rel family of transcription factors (NFκB family). Members of this family are activated by a variety of stimuli, including cytokines, growth factors, and oncogenic viruses. This gene encodes the p50 subunit of NFκB, which is a component of the NFκB transcription factor complex. (https://www.ncbi.nlm.nih.gov/gene/5966) |
| 3340    | NDST1         | N-deacetylase and N-          | This gene encodes a protein that belongs to the GcNAc-N-deacetylase/sulfotransferase family. The encoded enzyme is a type-II transmembrane protein that resides in the Golgi apparatus, where it catalyzes the transfer of sialic acid from 3-phosphoadenosine-5-phosphosulfate to N-glycolylneuraminic acid in heparan sulfate (https://www.ncbi.nlm.nih.gov/gene/3340) |
| ID   | ABBREVIATIONS | GENES                        | DESCRIPTION                                                                                                                                 |
|------|---------------|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 8880 | FUBP1         | Far upstream element-binding protein 1 | The protein encoded by this gene is a single stranded DNA-binding protein that binds to multiple DNA elements, including the far upstream element (FUSE) located upstream of c-myc. This protein is also thought to bind RNA and contains 3'-5' helicase activity with in vitro activity on both DNA-DNA and RNA-RNA duplexes. Aberrant expression of this gene has been found in malignant issues, and this gene is one of the most important of the helicase genes (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=8880). |
| 50486| G0S2          | G0/G1 switch 2               | This gene encodes a member of the G0/G1 family of matrix metalloproteinases (MMPs). Proteins in this family are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, cell migration, and tissue remodeling, as well as in disease processes such as atherosclerosis, tumor invasion, and removal. The encoded protein participates in the regulation of the mitotic cell cycle and S phase entry by cyclin-dependent protein kinase activities (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=50486). |
| 10765| KDM5B         | Lysine demethylase 5B        | This gene encodes a lysine-specific histone demethylase that belongs to the jumonji/ARID domain-containing family of histone demethylases. The protein encoded by this gene is a single stranded DNA-binding protein that binds to multiple DNA elements, including the far upstream element (FUSE) located upstream of c-myc. This protein is also thought to bind RNA and contains 3'-5' helicase activity with in vitro activity on both DNA-DNA and RNA-RNA duplexes. Aberrant expression of this gene has been found in malignant issues, and this gene is one of the most important of the helicase genes (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=10765). |
| 4312 | MMP1          | Matrix metalloproteinase 1   | The protein encoded by this gene is a nuclear receptor that is closely related to the estrogen receptor. This protein is involved in the binding of estradiol, a key sex hormone, to DNA and regulates the expression of genes involved in cell proliferation, differentiation, and apoptosis. The encoded protein participates in the regulation of the mitotic cell cycle and S phase entry by cyclin-dependent protein kinase activities (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=4312). |
| 94059| LENS9         | Leukocyte receptor cluster member 9 | This gene encodes a protein involved in signaling and organization of cell junctions during embryogenesis (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=94059). |
| 3643 | INSR          | Insulin receptor             | This gene encodes a member of the tyrosine kinase family of proteins. Binding of insulin or other ligands to this receptor activates the insulin signaling pathway, which regulates glucose uptake and release, as well as the synthesis and storage of carbohydrates, lipids, and protein (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=3643). |
| 8639 | AOC3          | Amine oxidase, copper containing 3 | This gene encodes a protein involved in signaling and organization of cell junctions during embryogenesis (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=8639). |
Figure 1. “Confidence view” of the protein-protein interaction network on String. The thickness of the blue line between 2 nodes indicates the level of confidence of association between these 2. The greater the thickness, the higher the level of confidence. The confidence score for each interaction indicates the prediction of a link between 2 nodes in the same metabolic pathway at the level of the KEGG database. The size differs between proteins of which the 3-dimensional structure (large sized) is known. Nodes whose 3-dimensional structure is not yet available are small sized.

Figure 2. Types of interactions between the network proteins on String.
Table 7. GO biological processes on String.

| BIOLOGICAL PROCESS (GO) | PATHWAY ID | PATHWAY DESCRIPTION | COUNT IN GENE SET | FALSE-DISCOVERY RATE |
|-------------------------|------------|---------------------|-------------------|----------------------|
| Mitochondrial ATP synthesis coupled proton transport | G0:0042776 | 7 | 3.03e–09 |
| Acute inflammatory response | G0:0002526 | 8 | 1.5e–07 |
| Regulation of cell death | G0:0010941 | 21 | 1.27e–06 |
| Signal transduction | G0:007165 | 36 | 1.43e–06 |
| Positive regulation of cellular metabolic process | G0:001325 | 28 | 1.52e–06 |
| Innate immune response | G0:0045087 | 17 | 1.52e–06 |
| Defense response | G0:005952 | 20 | 2.2e–06 |
| Response to stimulus | G0:005896 | 43 | 4.08e–06 |
| Positive regulation of NF kappaB transcription factor activity | G0:0051092 | 8 | 4.08e–06 |
| Single organism signaling | G0:004700 | 36 | 4.92e–06 |
| Cell surface receptor signaling pathway | G0:0007166 | 23 | 6.42e–06 |
| Regulation of apoptotic process | G0:0042981 | 19 | 7.02e–06 |
| Positive regulation of protein metabolic process | G0:0051247 | 19 | 7.57e–06 |
| Positive regulation of transcription, DNA-templated | G0:0045893 | 19 | 8.49e–06 |
| Positive regulation of cellular process | G0:0048522 | 33 | 9.52e–06 |

Abbreviation: GO, gene ontology.

Figure 3. Gene network customized by the network analyzer plugin on Cytoscape. The nodes of the network represent the proteins: each node represents all the proteins produced by a single gene locus encoding a protein. The edges represent protein-protein associations: the associations must be specific and significant, simply put, the proteins contribute together to a shared function; it does not necessarily mean that they physically bond to each other.

process (GO: 0045834);—response to organic substance (GO: 0010033).

Cytoscape results

The result network analyzer plugin for network visualization is shown in (Figure 3). The central genes for this network according to the chosen parameters are as follows: INS, IL1B, NFKBIA, and RELA, congruent with the results obtained by the String software.

The GO terms found by BiNGO pluging are displayed as a table of GO terms (Table 8). The functions are grouped into biological processes and the significant ones are as follows: positive regulation of lipid metabolic process, the multiorganism process, positive regulation of the cellular process, and response to an organic substance.

Positive regulation of lipid metabolic process. Of the 51 genes interacting in the network (Figure 3), 8 genes are annotated in this biological process according to the Table 8. This biological
| GO-ID | Description                                                                 | Cluster Frequency | Total Frequency | Genes                                                                 |
|-------|------------------------------------------------------------------------------|-------------------|----------------|----------------------------------------------------------------------|
| 42776 | Mitochondrial ATP synthesis coupled proton transport                          | 7/50 (14.0%)      | 12/14 (30.6%)  | ATP5B, ATP5A1, ATP5D, ATP5C1, ATP5H, ATP5O, ATP5B, ATP5A1, ATP5D,   |
|       |                                                                              |                   |                | ATP5C1, ATP5G3, ATP5H, ATP5O, ILA IRIS IL1B APOA1, PPARG, PPARA, VAY2, |
| 1470  | Positive regulation of lipid metabolic process                                | 7/50 (14.0%)      | 12/14 (30.6%)  | IL1A, IRS1, IL1B, ApOA1, ppARG, ppARA, vAv2, INS                     |
| 14985 | Response to organic cyclic substance                                          | 7/50 (14.0%)      | 12/14 (30.6%)  | IL1A, SOCS3, CDKN2B, SERPINA1, IL1B CASP1, RELA, TACR1 RELA, ICAM1, |
|       |                                                                              |                   |                | NFKBIA, IL1A, SOCS3, IL1B, RELA, . . .                              |
| 34220 | Ion transmembrane transport                                                   | 7/50 (14.0%)      | 12/14 (30.6%)  | ATP5B, ATP5A1, ATP5D, ATP5C1, ATP5H, ATP5O, ATP5B, ATP5A1, ATP5D, |
|       |                                                                              |                   |                | ATP5C1, ATP5G3, ATP5H, ATP5O, NFKBIA, IL1A, SOCS3, IL1B CASP1, . .   |
| 19216 | Response to organic substance                                                 | 7/50 (14.0%)      | 12/14 (30.6%)  | CDKN2B, SERPINA1, IRS1, PPARG, PPARA, VAY2, INS                      |

The table has displayed the most overrepresented GO terms, sorted by their P-value (ascending order from top to bottom). The board is a list of GO terms, with their names and GO-IDs, for the uncorrected P-value and the corrected P-value.
process has a significant \( P \) value of \( 5.2406 \times 10^{-9} \). The positive regulation of the lipid metabolic process (Figure 4) is related with other term children, the most significant are as follows: the positive regulation of the metabolic process of fatty acids, the positive regulation of the catabolic process of lipids, and the positive regulation of the lipid kinase activity.

**The multiorganism process.** Sixteen genes are annotated in this biological process according to Table 8. This biological process has a very significant \( P \) value of \( 4.9312 \times 10^{-9} \). The multiorganism process (Figure 5) is related with other term children, the most significant are as follows: the response to another organism and the interspecific interaction between organisms.

**Positive regulation of the cellular process.** This biological process has a significant \( P \) value of \( 8.6679 \times 10^{-9} \), with a ratio of 24 genes annotated (Table 8). The most significant term children from the positive regulation of the cellular process (Figure 6) are as follows: the positive regulation of the cellular communication, the positive regulation of the cellular metabolic process, and positive regulation of organelle organization.

**The response to an organic substance.** Sixteen genes are annotated in the response to an organic substance according to the Table 8. This biological process has a significant \( P \) value of \( 2.0898 \times 10^{-8} \). The most significant term children in this biological process are (Figure 7): The response to the cyclic organic substance, the response to molecules of bacterial origin, and response to the hormonal stimulus.

The BiNGO pluging ontological analysis revealed the involvement of functional network genes in biological processes related to metabolism, communication, and survival of epithelial cells of the gut of newborns. These results are in accordance with the results obtained by String software, showing the positive impact of active foods on the homeostasis of the newborns' intestines.

**Conclusion**

The results of coexpression and ontological studies provide insights into global patterns of gene expression in epithelial cells of term infants. The 5 central proteins in the networks (IL1β, RELA, INS, IRS1, and NFKBIA) are the major regulators of 4 significant biological processes. These biological processes induced in the first few months of a newborns' life have concerned intestinal development, effect of nutrition,
and impact of other environmental exposures on the intestinal microbiota colonization. Thus, this study offers a new depiction of the results to allow a better understanding of several interactions and their importance in health homeostasis.

**Contribution to the Field**
Breastfeeding is a strategy favored by evolution, to help our descendants survive and project our genes to succeeding generations. Thus, breast milk is a vector of bacteria in the days and months after birth. Gut microbiota established during these first months of life is vital for infant health and subsequent adults. This work aims to exploit the bioinformatics approaches, to provide a new representation and interpretation of the interactions between differentially expressed genes (DEGs) in the host intestine induced by the microbiota. The results of coexpression and ontological studies provide insights into global patterns of gene expression in epithelial cells of term infants. The significant biological processes induced by the central proteins in the networks have concerned intestinal development, effect of nutrition and impact of other environmental exposures on the intestinal microbiota colonization. So, this study can contribute to a new representation of complex interactions between microorganisms genes and host genome during the development of the intestine allowing a better understanding of several interactions and their importance in health homeostasis.

**REFERENCES**

1. Jost T, Lacroux C, Braegger CP, Rochat F, Chassard C. Vertical mother–neonate transfer of maternal gut bacteria via breastfeeding. *Environ Microbiol*. 2014;16:2891-2904. doi:10.1111/1462-2920.12238.
2. Kumar H, du Toit E, Kulkarni A, et al. Distinct patterns in human milk microbiota and fatty acid profiles across specific geographic locations. *Front Microbiol*. 2016;7:1619. doi:10.3389/fmicb.2016.01619.
3. Pannaraj PS, Li F, Cerini C, et al. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr*. 2017;171:647-654. doi:10.1001/jamapediatrics.2017.0378.
4. Voigt AY, Costea PI, Kühn TA, et al. Temporal and technical variability of human gut metagenomes. *Genome Biol*. 2015;16:73. doi:10.1186/s13059-015-0639-8.
5. Bashan A, Gibson TE, Friedman J, et al. Universality of human microbial dynamics. *Nature*. 2016;534:259-262. doi:10.1038/nature18301.
6. Caporaso JG, Lauber CL, Costello EK, et al. Moving pictures of the human microbiome. *Genome Biol*. 2011;12:R50. doi:10.1186/gb-2011-12-5-r50.
7. Flores GE, Caporaso JG, Henley JR, et al. Temporal variability is a personalized feature of the human microbiome. *Genome Biol*. 2014;15:531. doi:10.1186/s13059-014-0531-y.
8. Gründum MM, Gueimonde M, Laitinen K, et al. Maternal breast-milk and intestinal bifidobacteria guide the compositional development of the bifidobacterium microbiota in infants at risk of allergic disease. *Clin Exp Allergy*. 2007;37:1764-1772. doi:10.1111/j.1365-2222.2007.02849.x.
9. Bezirtzoglou E, Tsiotisias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*. 2011;17:478-482. doi:10.1016/j.anaerobe.2011.03.009.
10. Duijts L, Jaddoe VW, Hofman A, Moll HA. Prolonged and exclusive breastfeeding reduces the risk of infectious diseases in infancy. *Pediatrics*. 2010;126:e18-e25. doi:10.1542/peds.2008-3256.
11. Walker WA, Iyengar RS. Breast milk, microbiota, and intestinal immune homeostasis. *Pediatr Res*. 2015;77:220-228. doi:10.1038/pr.2014.160.
12. Bode L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*. 2012;22:1147-1162. doi:10.1093/glycob/cws074.
13. Chegani F. Effects of *Streptococcus thermophiles* bacteria on rat gene expression profiles. http://tesionline.unicatt.it/handle/10280/962. Updated February 24, 2011. Accessed May 2, 2019.
14. Mazmanian SK, Kasper DL. The love-hate relationship between bacterial polysaccharides and the host immune system. *Nat Rev Immunol*. 2006;6:849-858. doi:10.1038/nri1956.
15. Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol*. 2017;25:217-228. doi:10.1016/j.tim.2016.11.008.
16. Pappas KM, Louis E, Minton N, Mukhopadhyay B, Yang S. Genetic and Genome-Wide Insights into Microbes Studied for Bienergy. *Lausanne, Switzerland: Frontiers Media SA*; 2017.

**ORCID iD**
Badreddine Nouadi https://orcid.org/0000-0001-5175-4601
17. Schwartz S, Friedberg L, Ivanov IV, et al. A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biol*. 2012;13:r32. doi:10.1186/gb-2012-13-4-r32.

18. Knight JM, Davidson LA, Herman D, et al. Non-invasive analysis of intestinal development in preterm and term infants using RNA-sequencing. *Sci Rep*. 2014;4:5453. doi:10.1038/srep05453.

19. Kohan AB, Wang F, Lo C-M, Liu M, Tso P. ApoA-IV: current and emerging roles in intestinal lipid metabolism, glucose homeostasis, and satiety. *Am J Physiol Gastrointest Liver Physiol*. 2015;308:G472-G481. doi:10.1152/ajpgi.00998.2014.

20. Lukens J, Dixit VD, Kanneganti T-D. Inflammasome activation in obesity-related inflammatory diseases and autoimmunity. *Disco Med*. 2011;12:65-74.

21. Ferguson PJ, Laxer RM. New discoveries in CRMO: IL-1β, the neutrophil, and the microbiome implicated in disease pathogenesis in Pstpip2-deficient mice. *Semin Immunopathol*. 2015;37:407-412. doi:10.1007/s00281-015-0488-2.

22. Yang C-A, Chiang B-L. Inflammasomes and human autoimmunity: a comprehensive review. *J Autoimmun*. 2015;61:1-8. doi:10.1016/j.jaut.2015.05.001.

23. Lukens JR, Gurung P, Vogel P, et al. Dietary modulation of the microbiome affects autoinflammatory disease. *Nature*. 2014;516:246-249. doi:10.1038/nature13788.

24. Gurung P, Kanneganti T-D. Novel roles for caspase-8 in IL-1β and inflammasome regulation. *Am J Pathol*. 2015;185:17-23. doi:10.1016/j.ajpath.2014.08.025.

25. Oberst A, Dillon CP, Weinlich R, et al. Catalytic activity of the caspase-8-FLIPL complex inhibits RIPK3-dependent necrosis. *Nature*. 2011;471:363-367. doi:10.1038/nature09952.

26. Ben-Sasson SZ, Hu-Li J, Quiel J, et al. IL-1 acts directly on CD4 T cells to enhance their antigen-driven expansion and differentiation. *Proc Natl Acad Sci USA*. 2009;106:7119-7124. doi:10.1073/pnas.0902745106.

27. Seo DH, Che X, Kwak MS, et al. Interleukin-33 regulates intestinal inflammation by modulating macrophages in inflammatory bowel disease. *Sci Rep*. 2017;7:851. doi:10.1038/s41598-017-00840-2.

28. Al-Sadi R, Ye D, Suid HM, Ma-TY. IL-1β-induced increase in intestinal epithelial tight junction permeability is mediated by MEKK-1 activation of canonical NF-κB pathway. *Am J Pathol*. 2010;177:2310-2322. doi:10.2353/ajpath.2010.100371.

29. Steinbrecher KA, Harmel-Laws E, Sitcheran R, Baldwin AS. Loss of epithelial RelA results in deregulated intestinal proliferative/apoptotic homeostasis and susceptibility to inflammation. *J Immunol*. 2008;180:2588-2599.

30. Zhong XS, Winston JH, Luo X, et al. Neonatal colonic inflammation epigenetically aggravates epithelial inflammatory responses to injury in adult life. *Cell Mol Gastroenterol Hepatol*. 2018;6:65-78. doi:10.1016/j.mgh.2018.02.014.

31. Esposito DL, Ariu F, Lattanzio R, et al. The insulin receptor substrate 1 (31r1) intestinal epithelial differentiation and in colorectal cancer. *PLoS ONE*. 2012;7:e36190. doi:10.1371/journal.pone.0036190.

32. Hakuno F, Fukushima T, Yoneyama Y, et al. The novel functions of high-molecular-mass complexes containing insulin receptor substrates in mediation and modulation of insulin-like activities: emerging concept of diverse functions by IRS-associated proteins. *Front Endocrinol*. 2015;6:73. doi:10.3389/fendo.2015.00073.

33. Copp KD, White MF. Regulation of insulin sensitivity by serine/threonine phosphorylation of insulin receptor substrate proteins IRS1 and IRS2. *Diabetologia*. 2012;55:2565-2582. doi:10.1007/s00125-012-2644-8.

34. Shaw EJ, Smith EE, Whittingham-Dowd J, Hodges MD, Else KJ, Rigby RJ. Intestinal epithelial suppressor of cytokine signaling 3 (SOCS3) impacts on mucosal homeostasis in a model of chronic inflammation. *Immun Inflamm Dis*. 2017;5:336-345. doi:10.1002/iid3.171.

35. Thagia I, Shaw EJ, Smith E, Else KJ, Rigby RJ. Intestinal epithelial suppressor of cytokine signaling 3 enhances microbial-induced inflammatory tumor necrosis factor-α, contributing to epithelial barrier dysfunction. *Am J Physiol Gastrointest Liver Physiol*. 2015;308:G225-G31. doi:10.1152/ajpgi.00214.2014.

36. Khakh BS, Burnstock G. The double life of ATP. *Sci Am*. 2009;301:84-92.