Seeking Windows of Opportunity to Shape Lifelong Immune Health: A Network-Based Strategy to Predict and Prioritize Markers of Early Life Immune Modulation

Jolanda H. M. van Bilsen[1]*, Remon Dulos[1], Mariël F. van Stee[1], Marie Y. Meima[1],
Tanja Rouhani Rankouhi[1], Lotte Neergaard Jacobsen[2], Anne Staudt Kvistgaard[2],
Jossie A. Garthoff[3], Léon M. J. Knippels[4,5], Karen Knipping[4,5], Geert F. Houben[1],
Lars Verschuren[1], Marjolein Meijerink[1] and Shaji Krishnan[1]

[1] Netherlands Organisation for Applied Scientific Research (TNO), Zeist, Netherlands, [2] Arla Foods Ingredients, Aarhus, Denmark, [3] Danone Food Safety Center, Utrecht, Netherlands, [4] Danone Nutricia Research, Utrecht, Netherlands, [5] Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Utrecht, Netherlands

A healthy immune status is strongly conditioned during early life stages. Insights into the molecular drivers of early life immune development and function are prerequisite to identify strategies to enhance immune health. Even though several starting points for targeted immune modulation have been identified and are being developed into prophylactic or therapeutic approaches, there is no regulatory guidance on how to assess the risk and benefit balance of such interventions. Six early life immune causal networks, each compromising a different time period in early life (the 1st, 2nd, 3rd trimester of gestations, birth, newborn, and infant period), were generated. Thereto information was extracted and structured from early life literature using the automated text mining and machine learning tool: Integrated Network and Dynamical Reasoning Assembler (INDRA). The tool identified relevant entities (e.g., genes/proteins/metabolites/processes/diseases), extracted causal relationships among these entities, and assembled them into early life-immune causal networks. These causal early life immune networks were denoised using GeneMania, enriched with data from the gene-disease association database DisGeNET and Gene Ontology resource tools (GO/GO-SLIM), inferred missing relationships and added expert knowledge to generate information-dense early life immune networks. Analysis of the six early life immune networks by PageRank, not only confirmed the central role of the "commonly used immune markers" (e.g., chemokines, interleukins, *IFN, TNF, TGFB,* and other immune activation regulators (e.g., *CD55, FOXP3, GATA3, CD79A, C4BPA*), but also identified less obvious candidates (e.g., *CYP1A2, FOXK2, NELFCD, RENBP*). Comparison of the different early life periods resulted in the prediction of 11 key early life genes overlapping all early life periods (*TNF, IL6, IL10, CD4, FOXP3, IL4, NELFCD, CD79A, IL5, RENBP,* and *IFNG*), and also genes that were only described in certain early life period(s). Concluding, here we describe a network-based approach that provides a science-based

and systematical method to explore the functional development of the early life immune system through time. This systems approach aids the generation of a testing strategy for the safety and efficacy of early life immune modulation by predicting the key candidate markers during different phases of early life immune development.

INTRODUCTION

The first 1,000 days of life is a period of growth and development in which the foundations of lifelong immune homeostasis and microbial colonization are established in humans (1). Alterations during this period, due to environmental and host factors, are considered to be potential determinants of health-outcomes later in life (2–4). Therefore, risk reduction measures or immune health interventions during these stages of life may be most effective and efficient for improving health, increasing quality of life, and lowering costs to society due to immune related diseases and disorders.

When developing immune health interventions in early life, the regulatory authorities (EFSA, JECFA) stress the need to address the safety of such interventions. However, currently there is no regulatory guidance about how to assess the risk and benefit balance of such interventions. At the moment final safety confirmation comes from expensive and lengthy clinical follow up studies using a set of guidelines (5–7). Therefore, a need for a science-based system approach to assess the safety and benefit of nutritional immune interventions, with a special focus on early life is clear. With such an approach animal testing can be reduced, refined or replaced.

Key to understanding the potential of early life immunity to shape lifelong immune health is the concept of ontogeny—the immune system development from fetal life through adulthood. Previously, our group made an inventory and compared the maturation of the immune systems of human, mouse, rat, and mini pig, based predominantly on existing (from literature) and newly generated histologic data (8). Critical time windows of immune organ development were identified in human and the above mentioned experimental species. However, less is known about the functional time frames of the developing immune system in humans. This knowledge is crucial to identify factors that need to be considered for assessing the safety and efficacy of early life nutritional interventions and exposure.

As the immune system is an enormously complex system, it is crucial to obtain more understanding about the biological structures and processes to be able to improve human (immune) health. However, due to the enormous wealth of information available, it is extremely difficult to obtain a complete picture of the biological basis of immune related diseases and health. Individual researchers are often restricted to so called "knowledge pockets" (9) covering only a small fraction of all available knowledge, and that fractional information is spread through literature or various databases. This fragmentation of information clearly hampers our understanding of the molecular processes underlying human health and disease. In order to obtain a complete picture, data integration from different sources is required.

Systems immunology combined with bioinformatics can provide sufficient knowledge to identify factors to assess the safety and efficacy of early life nutritional interventions and exposure (10–12). Recent technological advances permit collection and storage of large datasets at molecular and cellular levels (genes, gene products, metabolic intermediates, macromolecules, cells). So far, most studies or research groups collected data sets from several—omics-platforms to understand the larger (systems) picture by putting the pieces together, mostly through association networks (e.g., Protein-Protein Interaction network). Association networks are static and undirected networks. They provide lesser information than a directed causal network. However, creation of system-wide causal networks from omics data is a task that is largely tedious, and not pragmatic. This is because the amount of data spanning the molecular changes in spatio-temporal space is too large to capture the system knowledge within causal network in sufficient detail. Nevertheless, the dynamics of the immune system are better understood and characterized with the use of causal networks. Our intention here is to create causal networks of the early life immune system in a comprehensive and pragmatic manner.

Here, we generated causal immune networks in early life from literature sources that correspond to the 1st, 2nd, 3rd trimester of gestation (resp. EG, LG, MG), birth, newborn and infant period as part of a bioinformatics workflow, which also included subsequent network enrichment steps to generate comprehensive causal early life immune networks. The network-based approach developed here, enabled us to elucidate different phases of early life immune development in a systematical way to predict and prioritize biological functions and genes associated with immune functioning in early life. Moreover, this systems approach aids the development of a science-based testing strategy for assessing the safety and efficacy of early life immune modulation by predicting the key candidate markers during different phases of early life immune development.

MATERIALS AND METHODS

Generation of the Basis of Early Life-Immune Networks Using Text Mining

The entire bioinformatics workflow to generate human early life networks is depicted in **Figure 1**. The first step was to select relevant manuscripts describing immune mechanisms in early life. An inventory of the available literature regarding 6 immune developmental periods [1st/2nd/3rd trimester of
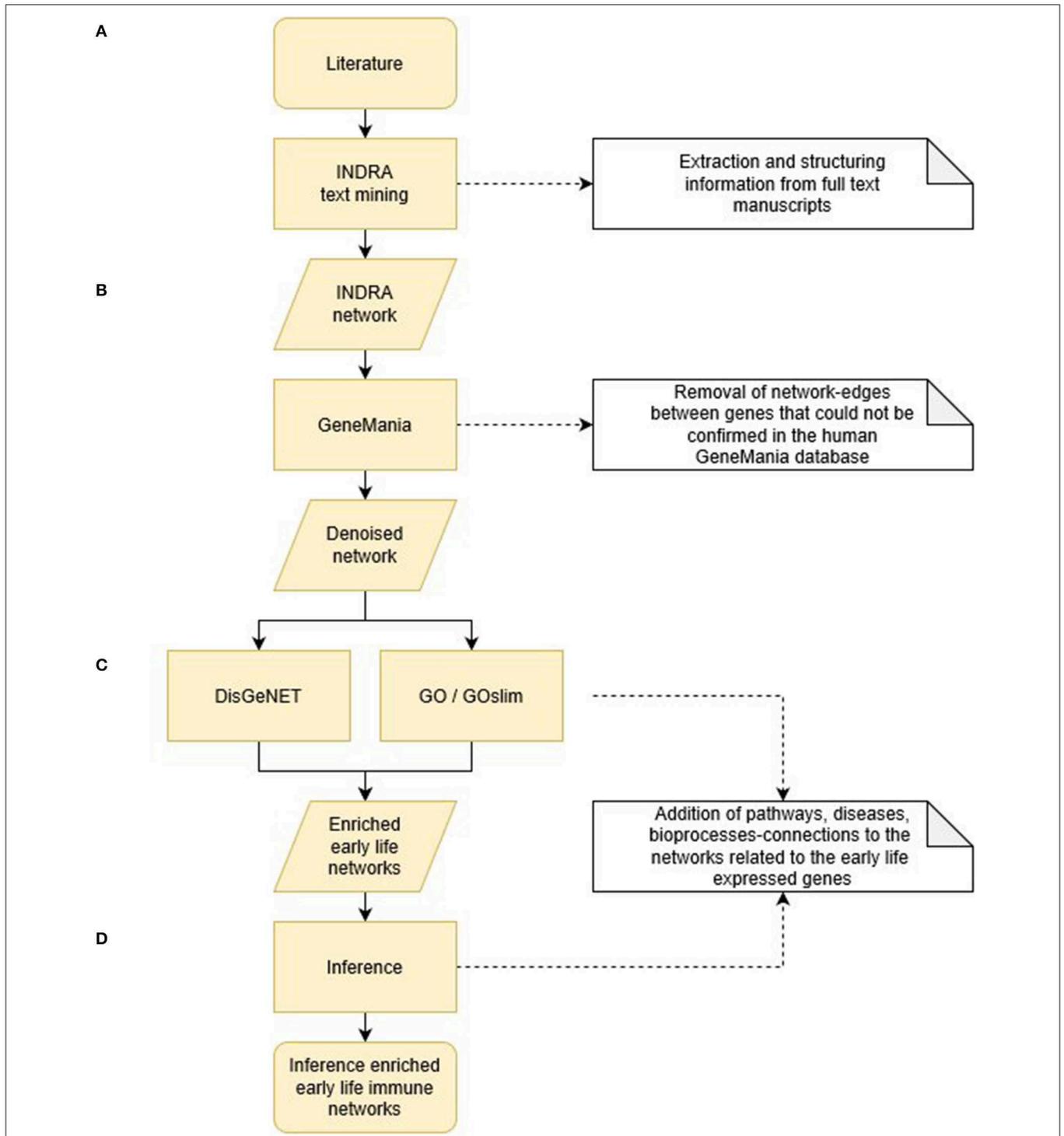
FIGURE 1 | Bioinformatics workflow to generate human early life networks. **(A)** Expert based selection of early life immune manuscripts were divided in 6 early life time periods and subjected to INDRA text mining tool. This resulted in 6 causal INDRA network. **(B)** The gene-gene connections of the INDRA networks were denoised and validated for the human situation by GeneMania. **(C)** DisGeNET and Gene Ontology tools (GO and GOslim) enriched the denoised early life networks by adding gene-disease connections and gene-process/pathway connections. **(D)** Inference calculations enriched the early life networks further by adding process-disease and disease-immune health endpoint connections. All steps together resulted in 6 human early life immune networks. The results of the different programming steps are depicted in **Tables 2–4** as indicated.

FIGURE 2 | Workflow to generate the basis of early life immune networks by literature. Six causal early life immune networks covering a different early life were generated by selecting appropriate manuscripts from literature after which relationships between biological entities were extracted by the text mining tool INDRA. Next INDRA assembled, de-duplicated and standardized all relationships into causal early life-immune networks each covering a different early life period. These INDRA networks formed the basis of the early life immune networks. *Several unique articles cover multiple early life periods.

gestation, birth, newborn (0–28 days), infant (1–24 months)] in human and experimental animals was made using Scopus and Medline (**Figure 2**). These databases were searched between 1st of December 2016 and 2nd of December 2016 and updated each half year (last update in March 2019). The search strings are depicted in **Table 1**. In total 2,966 articles were selected and manually screened on title, abstract and full text to select appropriate articles. Next, all selected articles were classified into the appropriate early life time period. The lengths of these different time periods in humans and experimental animals have been defined previously by Kuper et al. (8) and reported in **Table 2**.

The text from the manuscripts was moderately preprocessed to correct for obvious noise in text that interfered with the text analyses. Noise correction included deletion of special characters (except numbers, letters, punctuations and hyphens), "Materials

and Method" section, d.o.i., terms "fig." and "table," replacement of Greek characters by Roman letters, references containing "et al.," and hyphenation if a word was split into two parts at the end of a line of text. The Python code used to preprocess the manuscripts can be found at https://github.com/TNO/immune_health_textmining/blob/master/PDFminer.py.

After this preprocessing step, INDRA (Integrated Network and Dynamical Reasoning Assembler) text mining platform (www.indra.bio/) was used to extract relationships and structure information on causal mechanisms among biological entities from the selected articles. INDRA is an automated model assembly system interfacing with NLP systems and ontology databases to collect knowledge, and through a process of network assembly, produce causal graph and dynamical models (13–15).

INDRA text mining platform rendered the full texts of the selected articles computationally

TABLE 1 | Search strings used to assess the available literature regarding the immune functional developmental stages in human and experimental animals was performed by searching the databases Scopus and Medline.

Search terms	Combined with species terms	Combined with additional search terms
Thymus OR spleen OR lymph nodes OR Peyer's patches OR bone marrow OR liver	Human OR mini pig OR rat OR mouse	• Functional AND developmental AND stages OR • Immune AND development AND birth OR • Immune AND development AND weaning OR • Immune AND development AND prenatal OR • Immune AND development AND postnatal
Cord blood	Human OR mini pig OR rat OR mouse	• Functional AND developmental AND stages OR • Immune AND development AND birth OR • Immune AND development AND prenatal
*	Human OR mini pig OR rat OR mouse	• Functional AND developmental AND stages AND (amniotic fluid) OR placenta OR (*in utero*) OR intrauterine OR • Immune AND development AND (amniotic fluid) OR placenta OR (*in utero*) OR intrauterine AND birth OR • Immune AND development AND (amniotic fluid) OR placenta OR (*in utero*) OR intrauterine AND prenatal

No additional organ/tissue-specific term used in this search string which is specifically aimed at the gestational phase.

TABLE 2 | Developmental early life stages in human, minipig, rat, and mouse [adapted from (8)].

Early life period	EG[a]	MG	LG	Birth	Newborn	Infant
Human	GD0–GW12	GW13–28	GW29–40	–	0–28 days	1–23 months
Minipig	GD0–GD37	GD38–75	GD76–113	–	0–15 days	2–4 weeks
Rat	GD0–6	GD7–13	GD14–21	–	0–7/10 days	1/1.5–3 weeks
Mouse	GD0–6	GD7–13	GD14–21	–	0–7/10 days	1/1.5–3 weeks

[a]*Starts at fertilization/conception.*
EG/MG/LG, early/mid/late gestational period.
*GD, gestational day; GW, gestational week.*

accessible, identified biologically relevant entities (e.g., genes/proteins/metabolites/bioprocesses/diseases) and extracted relationships among these entities. Next, INDRA assembled, and standardized all relationships among the entities with associated evidence into causal early life-immune networks each covering a different early life period. Neo4J (https://Neo4j.com/) was used as a graph database management system to store, process and visualize the INDRA literature information as two-dimensional networks. This entire workflow is depicted in **Figure 2**.

Code used to generate the INDRA network is part of the INDRA repository and can be found at https://github.com/TNO/immune_health_textmining/blob/master/SRP_Neo4J.py.

Denoising INDRA Literature Networks
In order to eliminate noise from the INDRA literature networks and only depict those relationships for which there is a biological indication that the relationship is valid, all gene-gene relationships in the INDRA literature network were subjected to a denoising step using GeneMania (https://genemania.org/).

Genes coding for proteins described in the INDRA network were entered in the GeneMania Cytoscape plugin (freely available at http://genemania.org/plugin/) to identify human gene-gene associations from its large collection of organism specific functional association data that include protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity. These GeneMania-identified human gene-gene associations were compared to the gene-gene associations from the noisy INDRA literature networks, to identify and eliminate non-human specific associations between genes in the INDRA network. In the denoising step the edges (connections) between the genes were eliminated from the network, but not the genes themselves; they remained in the network as disconnected nodes. It must be noted that this step possibly eliminates true early-life gene-gene interactions if they are not represented in the human-specific GeneMania databases, which are mostly based on adult data. However, it is foreseen that this potential loss of information was compensated by the following enrichment steps because

FIGURE 3 | Overview of the steps used to enrich the INDRA networks. The genes described in early life literature (level 1). were entered in (i) DisGeNET to add gene-disease relationships to the network (level 1–4) and (ii) Gene Ontology tools GO/GO-SLIM to add gene-sub bioprocess (level 1–2), sub bioprocess—parent bioprocess (level 2–3) and gene-parent bioprocesses (level 1–3) relationships. Next the GO-terms linking to immune health features described previously in Meijerink et al. (16) were added to the network (level 2–5; blue arrow). The associations between bioprocesses and diseases (level 3–4) and disease–immune health features (level 4–5) were inferred (black arrows) based on the previous enrichment steps (orange arrows).

the disconnected genes remained part of the network. The code used to denoise the INDRA literature networks can be found at https://github.com/TNO/immune_health_textmining/blob/master/SRP_filter_networks.py.

Network Enrichments (Figure 3)

The INDRA network derived from literature reflects only the functionalities of the genes and processes described in literature which provides an incomplete picture of the functionalities of the described genes because the manuscripts usually focus on a specific topic. Therefore, it was important to determine whether the expressed genes are associated with a certain biological process and/or molecular function and/or diseases which were not addressed in the selected manuscripts. This knowledge was retrieved from several databases and added to the networks (enrichment). To enrich the INDRA early life immune literature networks, the genes coding for the proteins in the network were entered into the Gene Disease Association Database (DisGeNET; http://www.disgenet.org/) to retrieve the gene-disease associations using WebGestalt tool (17). The same sets of genes were also entered in the Gene Ontology resource tools (GO enrichment tools GO and GO-SLIM; http://geneontology.org/) to retrieve gene-bioprocess associations (GO/GO-SLIM).

As a final step in the network enrichments, the associations among bioprocesses, immune related diseases and immune health endpoints (16) were inferred based on the enrichment tool specific database knowledge of the number and similarity of the genes related to each of the network entities in different layers in the model (**Figure 3**). As described earlier, Neo4J (https://Neo4j.com/) was used as a graph database management system to store and process all network information, including the literature-derived information by INDRA.

Codes used to generate these enriched networks can be found at https://github.com/TNO/immune_health_textmining/blob/master/SRP_Neo4J.py https://github.com/TNO/immune_health_textmining/blob/master/SRP_add_endpoints_to_

disease_nodes.py and https://github.com/TNO/immune_health_textmining/blob/master/SRP_calc_inference.py.

Prioritization Immune Markers in Early Life

In order to identify key early life genes (hub genes), the PageRank centrality score was calculated in the early life networks. The PageRank analysis was launched by Google (the web search engine) to identify significant web pages (18–20) and has been used for the analysis of networks in identifying the important nodes in the network (21). Unlike simply calculating the connections of each gene in the network, the PageRank score measures the importance or popularity of a gene based solely on the interaction (link) structure of the interaction network. It selects the genes that exhibit a high degree, whilst also maintaining the important low-degree genes, which link to other important genes in the network. The underlying assumption is that more important genes are likely to receive more associations from other important genes/bioprocesses/diseases.

The PageRank algorithm code can be found at https://github.com/TNO/immune_health_textmining/blob/master/SRP_calc_pagerank_neo4j.py.

RESULTS

Generation of Early Life-Immune Literature Networks Using Text Mining

The literature covering the information on mechanisms involved in early life immune health is scattered across thousands of scientific papers. Therefore, text mining was applied to enable extracting and structuring information on causal mechanisms to create early life immune networks. In total 2,966 articles were selected using the search strings to explore literature databases. After manual screening 451 original manuscripts and 378 reviews were considered relevant (total number of selected 829 articles). This resulted in a selection of 249 articles for the

1st trimester of gestation, 296 articles for the 2nd trimester of gestation, 344 articles for the 3rd trimester of gestation, 252 articles for birth period, 287 articles for newborn period and 215 articles for the infant period. Please note that some articles covered multiple periods. From these full text articles, INDRA extracted resp. 2,101, 3,234, 3,654, 1,568, 2,917, and 1,487 unique relationships for the 1st, 2nd, 3rd trimester of gestation, birth, newborn and infant period (**Figure 2**). Next INDRA assembled,

FIGURE 4 | Early life immune networks based on information from early life immune literature and enriched with info from databases and inference steps, each covering a different phase during early life. **(A–C)** EG, MG, and LG; **(D)** birth; **(E)** newborn (0–28 days); **(F)** infant (1–24 months). **(G)** magnification of infant.
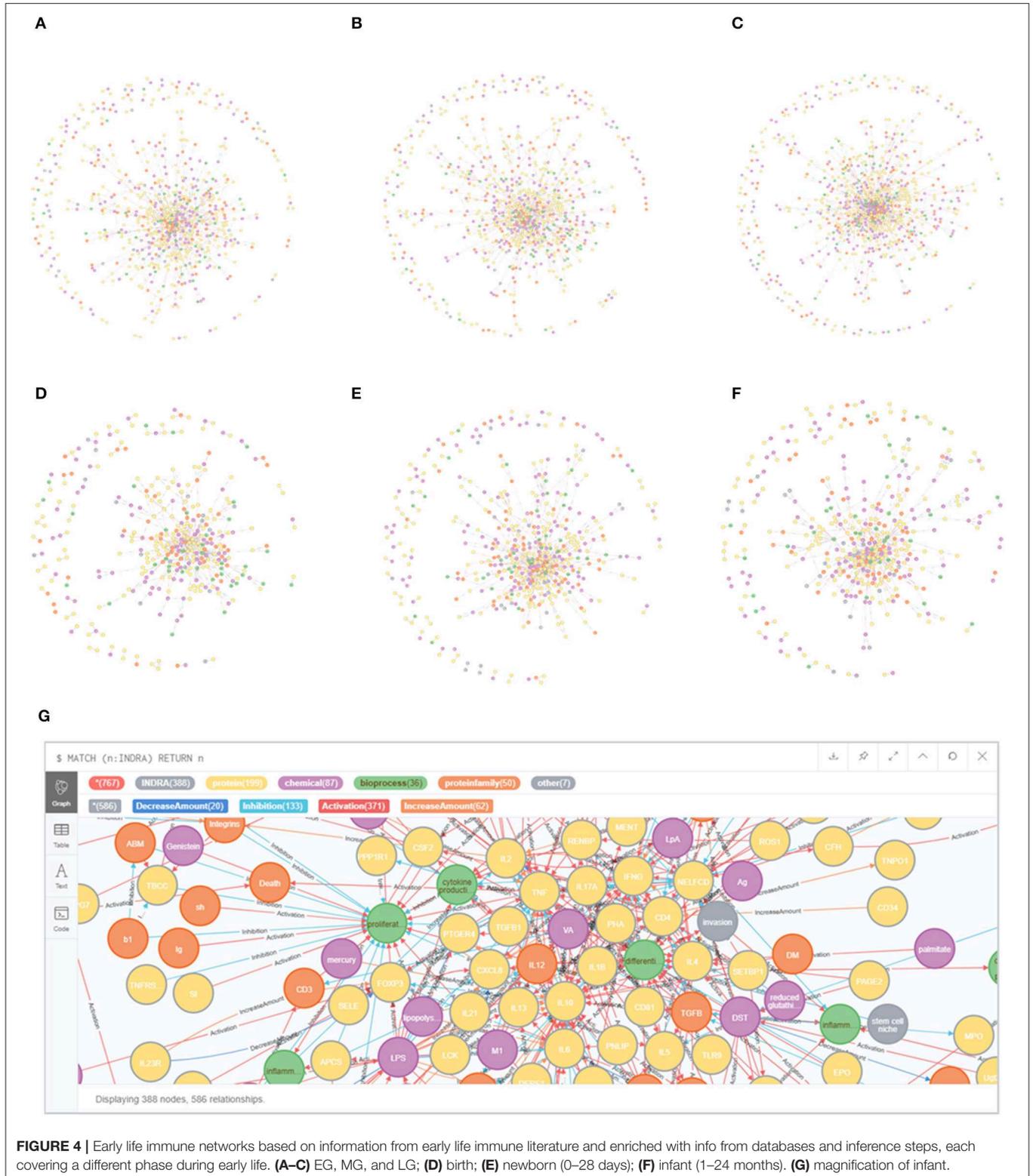
de-duplicated and standardized all relationships into 6 large early life-immune networks each covering a different early life period. The Neo4j-based framework enabled the visualization of the early life immune networks as depicted in **Figure 4**. As the networks are very dense in terms of numbers of nodes and edges, it is impossible to extract information directly from these networks without bioinformatical tools. The reason to depict these "unreadable" networks is to illustrate the complexity and density of them. In our methodology we identified 107 genes that have been described in the selected early life literature already during gestation and remained expressed throughout the infant period (**Supplementary Figure 1**).

Denoising Early Life-Immune Literature Networks (Table 3)

Approximately 30% (range 27–32%, depending on early life period) of the connections (edges) between the genes coding for proteins described in the INDRA network were overlapping with the human gene-gene interactions present in the GeneMania consulted databases (**Table 3**), indicating that the denoising step reduced ∼70% (depending on the early life network) of the

gene-gene connections in our network. This large reduction may be due to the fact that: (a) The gene-gene connection is solely relevant in early-life situations, which are not reflected in the GeneMania-consulted databases (which contain mainly adult data); (b) The gene-gene connection is non-human specific as the search strings for literature included guinea pig, rat, and mice; (c) Only genes that could be linked to a unique HUGO Gene Nomenclature Committee (HGNC) ID are recognized by GeneMania; and (d) The gene-gene connection is nonsense and should therefore be excluded. It must be noted that only the edges between the genes are removed, but the genes themselves remained part of the network. Although this elimination step possibly also eliminates some of the true early-life gene-gene interactions as suggested above, it is foreseen that this potential loss of information was compensated by the following enrichment steps.

Network Enrichments

The relationships of genes coding for the proteins that were identified in the early life networks by text mining were enriched by information retrieved from Gene Ontology and DisGeNET databases, respectively, is depicted in **Table 4**. After enrichment, the number of gene—bioprocess relationships were increased 60-fold (approximately). Of note, depending on the early-life time frame, DisGeNET databases introduced numerous gene-disease relationships (ranging from 1,719 to 4,568 relationships) to the early life immune networks. Other than this, the DisGeNET database not being specific to immune-related diseases, numerous non-immune diseases were also added to the early-life immune networks.

Subsequent addition of associations between bioprocesses and immune health endpoints (autoimmunity, hypersensitivity, resistance to neoplasms, resistance to infections) as previously described (16), further enriched the early life immune networks. As a final step in the network enrichments, the connections between bioprocesses and immune related diseases and immune health endpoints were inferred based on the knowledge of the number and the similarity of genes shared among the entities in different layers of the model (**Table 4** and **Figure 3**). The total number of nodes present in the early life immune networks

TABLE 3 | Number of edges between genes described in early life (literature info) and their presence in the human GeneMania database.

Network	#Genes/proteins* in early life literature extracted by text mining	#Gene-gene edges in early life literature	#Edges confirmed in GeneMania (%)
EG	440	228	72 (32%)
MG	477	278	84 (30%)
LG	508	319	90 (28%)
Birth	225	162	49 (30%)
Newborn	291	249	68 (27%)
Infant	232	174	51 (29%)

EG/MG/LG, early/mid/late gestation.
Sometimes it was not possible to distinguish protein names from corresponding gene names in literature. Therefore, all those names were annotated as being both a protein and a gene and regarded as 1 node in the network.

TABLE 4 | Results of enrichment/inference steps of the early life denoised INDRA immune networks.

	#Gene-bioprocess edges			#Gene-disease edges	#Bioprocess-immune endpoint edges	#Bioprocess-diseases edges	#Disease—immune endpoint edges
Source	Literature	GO-enrichment	GO-SLIM enrichment	DisGeNET enrichment	Meijerink et al. (16)	Inference	Inference
EG	149	9,546	443	3,894	1,121	1,701	1,023
MG	160	10,195	517	4,089	1,132	1,908	1,029
LG	180	10,968	546	4,568	1,246	2,207	1,136
Birth	67	3,929	168	1,719	695	1,073	627
Newborn	102	6,159	231	2,759	832	1,215	752
Infant	86	4,980	296	2,233	770	823	706

Depicted are the number of connections (edges) between biological entities (genes, bioprocesses, diseases, immune endpoints) added to the INDRA immune networks. EG/MG/LG, early/mid/late gestation.

TABLE 5 | Enriched early life immune network nodes.

Type of nodes	EG	MG	LG	Birth	Newborn	Infant
Proteins/genes*	440	477	508	225	291	232
Protein families	101	110	114	62	72	55
Chemicals	175	189	211	93	128	106
Bioprocesses**	51	56	58	36	39	34
GO processes	3,709	3,868	3,988	1,947	2,751	2,289
GOslim processes	55	55	55	59	60	59
Diseases	351	352	400	245	282	257
Immune health endpoint	4	4	4	4	4	4

*Depicted are the number of nodes in the networks after all enrichment/inference steps. These networks formed the basis of the gene prioritization (see **Table 6**). EG/MG/LG: early/mid/late gestation.*

*Using text mining, it was not always possible to distinguish genes from proteins (often same name used).

**Bioprocesses identified by ontology of INDRA text mining tool.

after the enrichment and inference steps are depicted in **Table 5**, indicating the complexity of the resulting 6 human early life immune networks.

Gene Prioritization to Identify Key Markers in Early Life

The enriched complex human early life immune networks formed the basis to identify the key markers in early life. The PageRank score of all nodes was calculated in the 6 human early life immune networks which resulted in 6 lists of prioritized immune markers each covering a different early life period (**Table 6**).

In general, the genes coding for the "commonly used immune markers" were highly ranked in all early life periods such as the cytokines including chemokines (e.g., *CXCL8, CXCL11, CXCL13*), interferons (*IFN*), interleukins (*IL1B, IL2, IL4, IL5, IL6, IL7, IL10, IL13, IL15, IL17A*), tumor necrosis factor (*TNF*), transforming growth *factor* (*TGFB*), and other immune activation regulators (e.g., *CD55, FOXP3, GATA3, CD79A, C4BPA*) directly involved in the immune response.

Comparison of the prioritized genes between the different early life periods (**Figures 5A,B**) showed that 36 genes were shown to be central in the network only during the gestational period, whereas others were more prominent in the periods birth, newborn and infant (6 genes: *RBP4, IL2, HAMP, env, ALG1,* and *IL1B*) or only in the infant period (14 genes: *TJP1, IL3, PIGS, ANPEP, CXCL11, CLCA3P, JAG1, NTAN1, CYYP1A2, CYP2E1, MADCAM1, VCAM1, GH1,* and *SCB*). Moreover, 11 genes were central in the early life immune networks covering all time periods: *TNF, IL6, IL10, CD4, FOXP3, IL4, NELFCD, CD79A, IL5, RENBP,* and *IFNG*. Most of these genes are immune related, however *RENBP*, renin binding protein, is an important regulator in the renin–angiotensin–aldosterone system. Moreover, *NELFCD*, Negative Elongation Factor Complex Member C/D, is an essential component of the NELF complex, which negatively regulates the elongation of transcription by RNA polymerase II.

Some of the top 50 genes were organ-specific such as *CPA1* (pancreas), *CRH* (neuronal), and *CDX2, MGAM,*

TABLE 6 | List of prioritized genes per early life time period.

Gene	EG	MG	LG	Birth	Newborn	Infant
CPA1	0.45	0.45	0.45			0.45
CA2	0.41	0.41	0.41			
MMP9	0.41	0.41	0.41			
SMO	0.39	0.39	0.39			
lectin	0.32	0.32	0.32			
PGR	0.30	0.30	0.30			
PSMA7	0.28	0.28	0.28			
TLR2	0.28	0.28	0.28			
PRC1	0.34	0.34	0.34			
DNAJB6	0.28	0.28		0.28		
CASP2	0.29	0.29				
ERVW-1	0.29	0.29				
lanA1	0.28	0.28	0.28			
ATJ1	0.28	0.28	0.28			
IL13	0.51	0.49	0.51	0.26	0.25	
CAV1	0.30	0.30	0.30			
MALL	0.28		0.28			
SPIC	0.28			0.28		
AR	0.34			0.28		
CPM	0.28					
MMP2	0.28					
PTGS2	0.28					
TNIP1	0.28					
CDX2	0.41	0.41	0.41	0.28		
ADCY1	0.52	0.60	0.52			
CXCL8	0.42	0.42	0.41	0.28	0.33	
MRLN	0.31	0.31	0.30			
CXCL13	0.35	0.36	0.29			
DSP	0.29	0.29	0.29			
FGF4	0.31	0.32		0.21		
SIRT1	0.30	0.35				
CD55	0.31	0.31	0.31			
CSH1	0.29	0.29	0.29			
cscK	0.30	0.29	0.30		0.30	
IL7	0.31		0.37			0.23
AHR	0.28			0.28	0.30	
C4BPA	0.29	0.29	0.29			
TNF	0.88	0.89	1.10	0.53	0.41	0.28
CD4	0.45	0.53	0.63	0.42	0.40	0.43
PAEP	0.35	0.35	0.35			
CRH	0.33	0.33	0.33			
MENT	0.28	0.30	0.30			
IFN	0.46	0.42	0.61	0.55		
IL6	0.61	0.67	0.75	0.46	0.70	0.49
IL5	0.29	0.29	0.61	0.27	0.25	0.66
FOXP3	0.33		0.50	0.62	0.65	0.47
IL10	0.53	0.76	0.72	0.98	0.79	0.97
CD79A	0.30	0.31	0.31	0.41	0.52	0.53
NELFCD	0.30	0.41	0.57	0.92	0.76	0.68
IL4	0.31	0.37	0.50	0.66	0.86	0.94
PREP		0.53	0.53			
TXN		0.28				0.28
ptc		0.28				
TEAD1		0.28				
TNFSF11		0.29				
TLR3		0.44	0.61			0.29
HLA-G		0.51	0.66			
LGALS1		0.28	0.34			
GALT		0.30	0.30			
RENBP		0.37	0.34	0.31	0.37	0.38
PTBP1			0.39	0.36	0.30	
DNTT			0.28		0.28	
MTOR			0.28			0.28
KLRK1			0.28			
GATA3			0.30	0.39	0.24	

Gene	EG	MG	LG	Birth	Newborn	Infant
CYP3A4			0.37		0.28	0.37
IFNG			0.33	0.47	0.36	0.41
ALG1				0.25	0.25	0.25
TGFB1				0.28	0.28	
Tnfrsf4				0.28	0.28	
ACKR2				0.25		
BCR				0.22		
CD19				0.22		
CREBBP				0.28		
EGR3				0.25		
GSR				0.21		
HPSE				0.21		
IL15				0.19		
IL17A				0.20		
lacZ				0.22		
NR3C2				0.21		
PTGER4				0.22		
RELB				0.24		
rpoD				0.22		
SLC14A1				0.28		
TBCC				0.27		
THPO				0.21		
RBP4				0.35	0.35	0.35
BTN3A1				0.24	0.25	
SET				0.21	0.24	
HAMP				0.29	0.28	0.40
LILRB4				0.31		0.38
LILRB2				0.21		0.28
env				0.25	0.27	0.26
IL2				0.30	0.34	0.45
IL1B				0.24	0.27	0.34
SGPL1					0.39	0.39
CDH1					0.28	0.28
CP					0.28	0.28
DCT					0.28	0.28
dop					0.28	0.28
FOXK2					0.28	0.28
TECR					0.28	0.28
ACAD8						0.28
CSF1R						0.28
GATA6						0.28
GC						0.28
HA					0.29	
HPSE						0.24
IRAK1						0.23
TLR4					0.29	
MGAM					0.23	0.36
SI					0.23	0.36
IGF1					0.23	0.28
TEC					0.26	0.28
TJP1						0.32
IL3						0.29
ANPEP						0.28
CLCA3P						0.28
CXCL11						0.28
JAG1						0.28
PIGS						0.28
NTAN1						0.25
CYP1A2						0.24
CYP2E1						0.24
MADCAM1						0.23
VCAM1						0.23
GH1						0.23
SCB						0.23

*The PageRank score of all nodes was calculated for each gene in order to identify the most "central" genes in the networks. The top 50 genes (i.e., highest PageRank score) per network are depicted, including their PageRank score. EG/MG/LG, early/mid/late gestation. Descriptions of the genes are described in **Supplementary Table 1**. The light to dark green-gradient reflects the increase in PageRank score.*

SI (intestine). Other genes were specifically involved in pregnancy such as *ERVW-1, CSH1, PAEP*, or involved in early life growth, and maturation (e.g., bone/cartilage *CA2,* cell cycle related proteins *CAV1, PRC1*; matrix modulation *FGF4, MMP9, MMP2*) were also identified as central markers.

Interestingly, also a few non-human genes were selected in the top 50 lists (*lectin, cscK, lacZ, rpoD, dop, AtJ1, lanA1, env, ptc*), representing plant, bacterial or viral specific proteins

FIGURE 5 | Venn diagram depicting unique and shared sets of genes from the top 50 gene lists of the different early life phases (**Table 6**); **(A)** number of genes and **(B)** gene names. For the gestational phases, the top 50 gene lists of early, mid and late period were combined, resulting in 67 unique genes. EG/MG/LG, early/mid/late gestation.
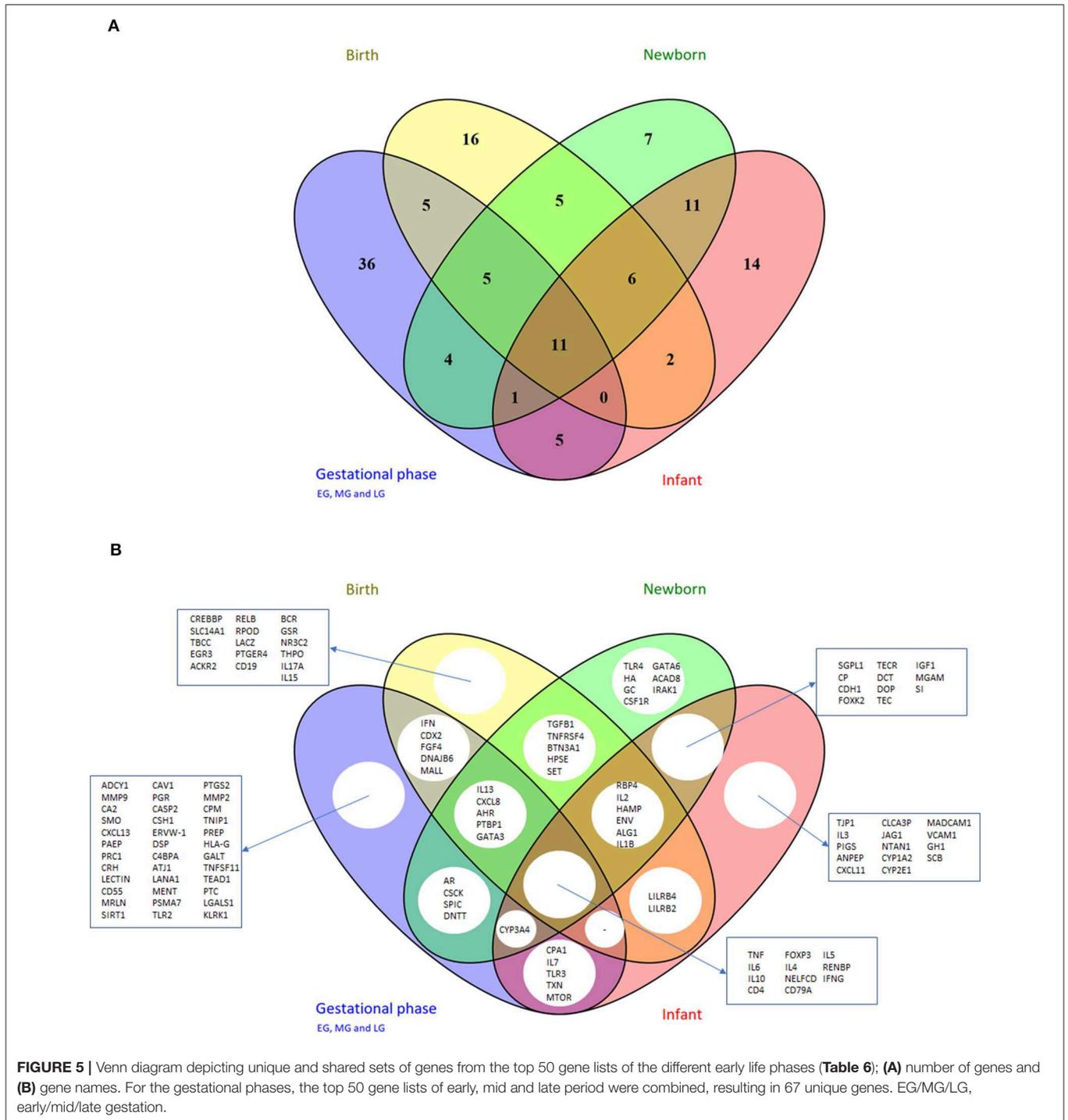
as key markers. So although the GeneMania denoising step eliminated the gene-gene edges of non-human genes, these non-human genes got central positions in the enriched early life networks.

Concluding, the PageRank analyses resulted in the identification of key early life genes with overlapping genes between the different early life periods, but also genes which were only described in a certain early life period. Moreover, the PageRank analyses confirmed the central role of the "commonly used immune markers" (cytokines, chemokines) in the early life networks, but also identified less obvious key marker candidates.

DISCUSSION

In this paper, we describe an approach to construct early life immune networks to identify and prioritize factors to assess safety and efficacy of early life immune modulation. As an alternative to expensive, hand-built models which can take months to years to construct, a workflow was created to generate causal early life immune networks. Literature-based interactions were used to form the basis of the network. These literature networks were denoised using GeneMania databases and enriched with data from comprehensive databases, such as Gene Ontology and DisGeNET. Thereafter, PageRank algorithm was applied to prioritize candidate genes in the early life networks. The entire pipeline is interpretable and intervenable in a way that domain experts can use our tools to greatly reduce the time required to identify relevant immune markers in early life.

Early life in humans is associated with large developmental milestones in the immune system.

Innate and adaptive immune cells are present early in the fetus during gestation and then expand significantly (8, 22). However, though the innate and adaptive immune cells are already present early during fetal development in the first trimester of gestation, the strength of their effector functions differ considerably from the adult situation. For instance, mature neutrophils are moderately present at the end of the first trimester, and increase steeply in number shortly before birth. Their number then returns to a stable level within days, but they show weak bactericidal functions, poor responses to inflammatory stimuli, reduced adhesion to endothelial cells and diminished chemotaxis (23).

Compared with the adult immune system, which has matured and evolved after years of exposure to antigens and environmental stimuli, the newborn immune system comes from a relatively sterile environment and is then rapidly exposed to microbial challenges (10). It is well-established that these differences in exposure to antigens and environmental stimuli have consequences when examining disease susceptibility. Severe infections remain a leading cause of neonatal morbidity and mortality. The immaturity of the immune system is thought to be an important factor for the increased rate of neonatal infections especially when born preterm but the basis for this is not fully understood (12), although the maturation of the neutrophil and endothelial adhesion function are thought to contribute significantly to the high risk of life-threatening infections in premature infants (23).

Many of our preventive strategies for neonates rely upon our understanding of the adult immune system, because of our limited knowledge of early life immunity. Therefore, there is no consensus regarding which factors should be covered to evaluate the safety and/or efficacy of the early life interventions and how all the available data should be interpreted appropriately. Our bioinformatics approach assumes that the functions of genes and proteins do not change over time. Instead, the biological balances between gene-sets expressed in early life and adult are assumed to change e.g., lower FOXP3 and CTLA-4 expression in activated regulatory T cells from human neonates compared to the adult situation (24). Therefore, the enrichment steps using information from databases (GO and DisGeNET) containing mostly data from adult situations, are assumed to be suitable to enrich the networks with functionalities of the genes/proteins that are described in early life literature. As input for these databases, only genes shown to be expressed in a specific early life period were entered to exclude the possibility that genes/proteins that are not (yet) expressed in that specific time frame would be introduced in the network. As others, we suggest that not the gene function as such, but the context in which the genes are expressed in early life determines the impact of the gene expression on the biological processes, cellular responses and/or cellular phenotype of the immune cell. Especially the microbial context has been suggested to be important: the interactions between the developing immune system and the microbes colonizing the intestine, skin and airways of a newborn child has been suggested by several groups (11, 25, 26). Olin et al. (11) showed that the microbiome diversity increased after birth but children with exceptionally lower diversity indicating bacterial dysbiosis (and high level of activated T cell populations) showed an increased immunological heterogeneity at 3 months of age. Several key immune cell populations (DCs, B cells, NK cells), reach adult-like phenotypes during the first 3 months of life, which suggests that environmental exposures during this period could have influence later in life. For example, differential susceptibility to autoimmunity and asthma may relate to DC exposure to bacterial antigens early in life, which could lead to more tolerogenic DCs later in life (27–29).

Currently only a few biomarkers of inflammation have been developed into biomarker assays approved and recommended by regulatory bodies for use in clinical studies, which includes CRP, TNF-α, serotransferrin and erythrocyte sedimentation rate (30). Although many candidate markers are identified based on preclinical and clinical studies (as listed in the Thompson Reuters Integrity[SM] Biomarkers Database), only a few are further validated and used for assay development highlighting the classical to clinical biomarkers gap. Moreover, in early life the identification of suitable markers is even more limited due to the fact that immunological studies on newborns tend to be small-scale and focus only on few factors because of limited sample volumes and low-throughput techniques as noted by Schaffert at al. (10). The early life immune networks generated in our approach enabled us to identify and rank genes that have the most central role in the early life immune networks. This is in contrast to earlier identified candidate markers for (pre-) clinical studies which are not specifically aimed at early life and not necessarily prioritized in a biological context.

There are multiple ways to prioritize genes in a biological network (31, 32). In computing network scores, most of the current approaches yield the limitation that the full network topology (systems approach) is not taken into account. Instead, such scoring methods focus on direct links or the most direct paths (shortest paths) within a constrained neighborhood around genes, ignoring potentially informative indirect paths. By applying PageRank algorithm, the full topology of the immune networks is taken into account.

Comparing the top 50 genes of the early life networks of the different time frames shows that many genes are already described in literature early in gestation. In general, the genes coding for the "commonly used immune markers" were highly

ranked in all early life periods such as the cytokines including chemokines and other immune activation regulators directly involved in the immune response. Interestingly, transcription factors GATA-3 and FOXP3 that regulate Th2 and T regulatory cell development are highly ranked in the networks, whereas the gene coding for T-bet (*TBX21*), the transcription factor for Th1 differentiation, was in the lower regions of the priority lists. It has been shown that these 3 transcription factors cross regulate one another: T-bet modulates GATA-3 function and Th2 cytokines block Th1 differentiation (33–36). Additionally, GATA-3 has been shown to inhibit FOXP3 transcription by binding to the FOXP3 gene promoter (37). The low priority ranking of the gene coding for T-bet is in line with the current view of an unbalanced Th1/Th2 neonatal immunity resulting in skewing toward Th2 immunity. Moreover, the genes related to Th17 responses [transcription factor gene coding for RORγT (*RORC*) and *IL17A, IL17F*, and *IL22*], are also of low priority (not in top 50) in the networks. In the context of the neonatal Th2-biased immune response, the inhibitory effect of IL-4 on the development of inflammatory Th17-type responses has been described to represent a major regulation mechanism (38) which may explain the low priority of Th17 related genes and the high priority of IL-4 in the early life networks.

Several non-human genes (lanA1, cscK, dop, rpoD, lacZ, env, ptc, lectin) were ranked in the top 50, which might seem unexpected or perhaps even suggest a flaw in the bioinformatics approach. However, their presence and relevance may well-explained. In our workflow in the denoising step using GeneMania, we removed the connections between genes that were not of human origins, but we did not exclude the non-human genes from the early immune networks: the non-human genes remained in the network as disconnected nodes.

The next step in the generation of early life immune networks was the addition of connections (edges) between the human and non-human genes to human pathways/diseases/bioprocesses (input DisGeNet and GO databases). Genes from rat/mouse/guinea pig will likely not be connected to human processes, so these genes will stay disconnected to the network and therefore have a very low priority in the PageRank scoring. However, some of the non-human genes from mainly viral or microbial origin could be connected in our workflow to multiple human processes/diseases and therefore turned out to be in the top 50 of the PageRank scoring. The relevance of the role of these non-human genes in immune responses could be confirmed by literature: lanA1 (viral protein LanA1; role in host-virus interaction) (39), cscK (bacterial fructokinase; role in TLR4 activation) (40), dop (bacterial pup deamidase; role in resistance to infection) (41), rpoD (bacterial sigma factor for RNA polymerase; role in exponential growth bacteria) (42), lectin (role in activation of innate immune system) (43), lacZ (bacterial beta-galactosidase; Th1-associated) (44), env (viral envelope glycoprotein gp160; role in immune evasion) (45).

Several genes, which are usually not regarded as immune-related, got a prominent position in our early life immune networks such as genes involved in pregnancy, growth, and maturation (e.g., *ERVW-1, CSH1, PAEP, CA2, CAV1, PRC1, FGF4, MMP9, MMP2*). Several intestinal digestion related genes (*MGAM, ANPEP, SI*) were present in the top 50 in the birth-newborn-infant networks, which might be related to start of oral diet after birth. These examples emphasize the role of the immune system on so many other non-immune bioprocesses, which should be taken into account during assessment of possible (side-)effects of immune modulation in early life. Indeed, several chemokines and cytokines selected in our workflow, such as CXCL8, IL-10, TNF, IL1B, TGFb are multifunctional molecules initially described as having a role in endometrial functions and play a role in appropriate embryo implantation or placental functioning (46, 47). Moreover, TNF and TGFb have been identified as core activators of epithelial to mesenchymal transition, which is essential for embryonic development (48, 49). Although our approach to collect and structure and prioritize all available information from literature and databases to identify candidate markers is exhaustive, it also has its limitations due to the natural limitations in the curation process of the usage of enrichment tool-dependent auxiliary databases, and to inaccuracies derived from text mining. Others being annotation issues, such as the incomplete annotation of genes to GO terms and diseases (50, 51). Furthermore, the approach might be subjected to a reporting bias as it can be difficult to distinguish the absence of a gene in early life or a relationship between molecules/pathways from a lack of evaluation. In addition, we do not take the context of the gene expression into account whereas it is known that the context determines greatly the impact of the genes on biological processes, cellular responses and/or cellular phenotypes of the immune cells. Also, the networks are not organ-specific, although organ-specific genes are in the top 50 of prioritized genes, such as *CPA1* (pancreas), *CRH* (brain), and *CDX2, MGAM, SI* (intestines).

The strength/weight of the relationships in the network were not taken into consideration, but merely 6 association networks have been generated of possible biological relationships in early life immunity. The next important step for the applicability of this approach would be to validate these relationships based on gene expression data, which will guide us to validate the networks and moreover enable us to finetune the weighing of the various relationships in the network. This may result in a re-prioritization of the most important genes in a specific period in early life. Moreover, by using gene expression data, it becomes possible to identify critical time frames for specific immune modulation, because depending on the exposure, different pathways/processes may be activated. Even taking into account these current limitations, to the best of our knowledge, this is the first global overview of the early life immune system that can be used as a starting point to select putative markers to monitor the functioning of the early life immune system.

The future step would be to enrich the early life immune networks with early life gene-expression data to generate a quantitative early life immune network for (i) the analysis of mechanisms underlying immune health and disease in early life and (ii) the validation of candidate markers of disease and health.

In conclusion, we describe a network-based approach that provides a science-based and systematic method to explore the functional development of the early life immune system in time. This systems approach aids the generation of a testing strategy for

assessing the safety and efficacy of early life immune modulation by predicting the key candidate markers during different phases of early life immune development.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JB, MM, LV, and SK contributed to the conception and design of the study. MM, MYM, and TR performed literature database searches and selection. RD, MS, and SK wrote scripts for the preprocessing of the manuscripts, GeneMania denoising, GO/DisGeNET database-searches and inference steps and PageRank algorithm score calculation. JB, MM, and SK wrote the manuscript. JB, RD, MS, MYM, TR, LJ, AK, JG, LK, KK, GH, LV, MM, and SK contributed to manuscript revision, read and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.00644/full#supplementary-material

Supplementary Figure 1 | Venn diagram depicting unique and shared sets of genes of the different early life phases extracted from literature by INDRA text mining (**Table 3**) without PageRank prioritization step. EG/MG/LG, early/mid/late gestation.

Supplementary Table 1 | Descriptions of the prioritized genes listed in **Table 6**.

REFERENCES

1. Dzidic M, Boix-Amorós A, Selma-Royo M, Mira A, Collado M. Gut microbiota and mucosal immunity in the neonate. *Med Sci*. (2018) 6:56. doi: 10.3390/medsci6030056

2. Baird J, Jacob C, Barker M, Fall C, Hanson M, Harvey N, et al. Developmental origins of health and disease: a lifecourse approach to the prevention of non-communicable diseases. *Healthcare*. (2017) 5:E14. doi: 10.3390/healthcare5010014

3. Hanson MA, Gluckman PD. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiol Rev*. (2014) 94:1027–76. doi: 10.1152/physrev.00029.2013

4. Hanley B, Dijane J, Fewtrell M, Grynberg A, Hummel S, Junien C, et al. Metabolic imprinting, programming and epigenetics - a review of present priorities and future opportunities. *Br J Nutr*. (2010) 104:S1–25. doi: 10.1017/S0007114510003338

5. Turck D, Bresson J, Burlingame B, Dean T, Fairweather-Tait S, Heinonen M, et al. Guidance on the preparation and presentation of an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283. *EFSA J*. (2016) 14:1–24. doi: 10.2903/j.efsa.2016.4594

6. Aguilar F, Crebelli R, Dusemund B, Galtier P, Gilbert J, Gott D, et al. Guidance for submission for food additive evaluations. *EFSA J*. (2012) 10:1–60. doi: 10.2903/j.efsa.2012.2760

7. Administrative guidance on the submission of applications for authorisation of a novel food pursuant to Article 10 of Regulation (EU) 2015/2283. *EFSA Support Publ*. (2018) 15:1–22. doi: 10.2903/sp.efsa.2018.EN-1381

8. Kuper CF, van Bilsen J, Cnossen H, Houben G, Garthoff J, Wolterbeek A. Development of immune organs and functioning in humans and test animals: implications for immune intervention studies. *Reprod Toxicol*. (2016) 64:180–90. doi: 10.1016/j.reprotox.2016.06.002

9. Cokol M, Iossifov I, Weinreb C, Rzhetsky A. Emergent behavior of growing knowledge about molecular interactions. *Nat Biotechnol*. (2005) 23:1243–47. doi: 10.1038/nbt1005-1243

10. Schaffert S, Khatri P. Early life immunity in the era of systems biology: understanding development and disease. *Genome Med*. (2018) 10:1–3. doi: 10.1186/s13073-018-0599-1

11. Olin A, Henckel E, Chen Y, Olin A, Henckel E, Chen Y, et al. Stereotypic immune system development in newborn children article stereotypic immune system development in newborn children. *Cell*. (2018) 174:1277–92.e14. doi: 10.1016/j.cell.2018.06.045

12. Kollmann TR, Kampmann B, Mazmanian SK, Marchant A, Levy O. Protecting the newborn and young infant from infectious diseases: lessons from immune ontogeny. *Immunity*. (2017) 46:350–63. doi: 10.1016/j.immuni.2017.03.009

13. Gyori BM, Bachman JA, Subramanian K, Muhlich JL, Galescu L, Sorger PK. From word models to executable models of signaling networks using automated assembly. *Mol Syst Biol*. (2017) 13:954. doi: 10.15252/msb.20177651

14. Duong D, Stone N, Goertzel B, Venuto J. Indra: emergent ontologies from text for feeding data to simulations. In: *Spring Simul Interoperability Work 2010, 2010 Spring SIW*. Monterey, CA. (2010). p. 385–94.

15. Sales JE, Souza L, Barzegar S, Davis B, Freitas A, Handschuh S. Indra: a word embedding and semantic relatedness server. In: *Lr 2018 - 11th International Conference on Language Resourses and Evaluation* (Miyazaki). (2019). p. 132–32.

16. Meijerink M, van den Broek T, Dulos R, Neergaard Jacobsen L, Staudt Kvistgaard A, Garthoff J, et al. The impact of immune interventions: a systems biology strategy for predicting adverse and beneficial immune effects. *Front Immunol*. (2019) 10:231. doi: 10.3389/fimmu.2019.00231

17. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res*. (2013) 41:W77–83. doi: 10.1093/nar/gkt439

18. Page L, Brin S, Motwani R, Winograd T. *The PageRank Citation Ranking: Bringing Order to the Web*. Stanford InfoLab (1999). Available online at: https://storm.cis.fordham.edu/~gweiss/selected-papers/classic-pagerank-paper.pdf

19. Dellavalle RP, Schilling LM, Rodriguez MA, Van de Sompel H, Bollen J. Refining dermatology journal impact factors using PageRank. *J Am Acad Dermatol*. (2007) 57:116–9. doi: 10.1016/j.jaad.2007.03.005

20. Griffiths TL, Steyvers M, Firl A. Google and the mind. *Psychol Sci*. (2007) 18:1069–76. doi: 10.1111/j.1467-9280.2007.02027.x

21. Bánky D, Iván G, Grolmusz V. Equal opportunity for low-degree network nodes: a PageRank-based method for protein target identification in metabolic graphs. *PLoS ONE*. (2013) 8:e54204. doi: 10.1371/journal.pone.0054204

22. Gollwitzer ES, Marsland BJ. Impact of early-life exposures on immune maturation and susceptibility to disease. *Trends Immunol*. (2015) 36:684–96. doi: 10.1016/j.it.2015.09.009

23. Nussbaum C, Gloning A, Pruenster M, Frommhold D, Bierschenk S, Genzel-Boroviczeny O, et al. Neutrophil and endothelial adhesive function during human fetal ontogeny. *J Leukoc Biol*. (2013) 93:175–84. doi: 10.1189/jlb.0912468

24. Rueda CM, Moreno-Fernandez ME, Jackson CM, Kallapur SG, Jobe AH, Chougnet CA. Neonatal regulatory T cells have reduced capacity

to suppress dendritic cell function. *Eur J Immunol*. (2015) 45:2582–92. doi: 10.1002/eji.201445371

25. Laforest-Lapointe I, Arrieta M-C. Patterns of early-life gut microbial colonization during human immune development: an ecological perspective. *Front Immunol*. (2017) 8:788. doi: 10.3389/fimmu.2017.00788

26. Vatanen T, Kostic AD, D'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell*. (2016) 165:842–53. doi: 10.1016/j.cell.2016.04.007

27. Pezoldt J, Pasztoi M, Zou M, Wiechers C, Beckstette M, Thierry GR, et al. Neonatally imprinted stromal cell subsets induce tolerogenic dendritic cells in mesenteric lymph nodes. *Nat Commun*. (2018) 9:3903. doi: 10.1038/s41467-018-06423-7

28. Daley D. The evolution of the hygiene hypothesis. *Curr Opin Allergy Clin Immunol*. (2014) 14:390–6. doi: 10.1097/ACI.0000000000000101

29. Domogalla MP, Rostan PV, Raker VK, Steinbrink K. Tolerance through education: how tolerogenic dendritic cells shape immunity. *Front Immunol*. (2017) 8:1764. doi: 10.3389/fimmu.2017.01764

30. Minihane AM, Vinoy S, Russell WR, Baka A, Roche HM, Tuohy KM, et al. Low-grade inflammation, diet composition and health: current research evidence and its translation. *Br J Nutr*. (2015) 114:999–1012. doi: 10.1017/S0007114515002093

31. Gonçalves JP, Francisco AP, Moreau Y, Madeira SC. Interactogeneous: disease gene prioritization using heterogeneous networks and full topology scores. *PLoS ONE*. (2012) 7:e49634. doi: 10.1371/journal.pone.0049634

32. Gill N, Singh S, Aseri TC. Computational disease gene prioritization: an appraisal. *J Comput Biol*. (2014) 21:456–65. doi: 10.1089/cmb.2013.0158

33. Hwang ES, Szabo SJ, Schwartzberg PL, Glimcher LH. T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. *Science*. (2005) 307:430–3. doi: 10.1126/science.1103336

34. Agnello D, Lankford CSR, Bream J, Morinobu A, Gadina M, O'Shea JJ, et al. Cytokines and transcription factors that regulate T helper cell differentiation: new players and new insights. *J Clin Immunol*. (2003) 23:147–61. doi: 10.1023/a:1023381027062

35. Messi M, Giacchetto I, Nagata K, Lanzavecchia A, Natoli G, Sallusto F. Memory and flexibility of cytokine gene expression as separable properties of human T(H)1 and T(H)2 lymphocytes. *Nat Immunol*. (2003) 4:78–86. doi: 10.1038/ni872

36. Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell*. (2000) 100:655–69. doi: 10.1016/S0092-8674(00)80702-3

37. Zheng W, Flavell RA. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell*. (1997) 89:587–96. doi: 10.1016/S0092-8674(00)80240-8

38. Weaver CT, Hatton RD, Mangan PR, Harrington LE. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Annu Rev Immunol*. (2007) 25:821–52. doi: 10.1146/annurev.immunol.25.022106.141557

39. Sun R, Liang D, Gao Y, Lan K. Kaposi's Sarcoma-associated herpesvirus-encoded LANA interacts with host KAP1 to facilitate establishment of viral latency. *J Virol*. (2014) 88:7331–44. doi: 10.1128/JVI.00596-14

40. Spruss A, Kanuri G, Wagnerberger S, Haub S, Bischoff SC, Bergheim I. Toll-like receptor 4 is involved in the development of fructose-induced hepatic steatosis in mice. *Hepatology*. (2009) 50:1094–104. doi: 10.1002/hep.23122

41. Cerda-Maira FA, Pearce MJ, Fuortes M, Bishai WR, Hubbard SR, Darwin KH. Molecular analysis of the prokaryotic ubiquitin-like protein (Pup) conjugation

pathway in *Mycobacterium tuberculosis*. *Mol Microbiol*. (2010) 77:1123–35. doi: 10.1111/j.1365-2958.2010.07276.x

42. Jishage M, Iwata A, Ueda S, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol*. (1996) 178:5447–51. doi: 10.1128/JB.178.18.5447-5451.1996

43. Brown GD, Willment JA, Whitehead L. C-type lectins in immunity and homeostasis. *Nat Rev Immunol*. (2018) 18:374–89. doi: 10.1038/s41577-018-0004-8

44. Ménoret S, Aubert D, Tesson L, Braudeau C, Pichard V, Ferry N, et al. lacZ transgenic rats tolerant for β -galactosidase: recipients for gene transfer studies using lacZ as a reporter gene. *Hum Gene Ther*. (2002) 13:1383–90. doi: 10.1089/104303402760128603

45. Cook JD, Lee JE. The secret life of viral entry glycoproteins: moonlighting in immune evasion. *PLoS Pathog*. (2013) 9:e1003258. doi: 10.1371/journal.ppat.1003258

46. Du MR, Wang SC, Li DJ. The integrative roles of chemokines at the maternal-fetal interface in early pregnancy. *Cell Mol Immunol*. (2014) 11:438–48. doi: 10.1038/cmi.2014.68

47. Salama KM, Alloush MK, Al hussini RM. Are the cytokines TNF alpha and IL 1Beta early predictors of embryo implantation? Cross sectional study. *J Reprod Immunol*. (2020) 137:102618. doi: 10.1016/j.jri.2019.102618

48. Xu W, Yang Z, Lu N. A new role for the PI3K/Akt signaling pathway in the epithelial-mesenchymal transition. *Cell Adhes Migr*. (2015) 9:317–24. doi: 10.1080/19336918.2015.1016686

49. Shook D, Keller R. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev*. (2003) 120:1351–83. doi: 10.1016/j.mod.2003.06.005

50. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol*. (2009) 5:290. doi: 10.1038/msb.2009.47

51. Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PLoS ONE*. (2011) 6:e20284. doi: 10.1371/journal.pone.0020284