

## Article

# Network Approaches to Integrate Analyses of Genetics and Metabolomics Data with Applications to Fetal Programming Studies

Alan Kuang <sup>1</sup>, M. Geoffrey Hayes <sup>2</sup>, Marie-France Hivert <sup>3,4,5</sup>, Raji Balasubramanian <sup>6</sup>, William L. Lowe, Jr. <sup>2</sup> and Denise M. Scholtens <sup>1,\*</sup>

- <sup>1</sup> Department of Preventive Medicine (Biostatistics), Northwestern University Feinberg School of Medicine, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL 60611, USA; alan.kuang@northwestern.edu
- <sup>2</sup> Department of Medicine, Northwestern University Feinberg School of Medicine, Rubloff 12, 420 E. Superior St, Chicago, IL 60611, USA; ghayes@northwestern.edu (M.G.H.); wlowe@northwestern.edu (W.L.L.J.)
- <sup>3</sup> Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA 02215, USA; mhivert@partners.org
- <sup>4</sup> Diabetes Unit, Endocrine Division, Massachusetts General Hospital, Boston, MA 02114, USA
- <sup>5</sup> Department of Medicine, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada
- <sup>6</sup> Department of Biostatistics and Epidemiology, University of Massachusetts-Amherst, Amherst, MA 01003, USA; rbalasub@schoolph.umass.edu
- \* Correspondence: dscholtens@northwestern.edu



**Citation:** Kuang, A.; Hayes, M.G.; Hivert, M.-F.; Balasubramanian, R.; Lowe, W.L., Jr.; Scholtens, D.M. Network Approaches to Integrate Analyses of Genetics and Metabolomics Data with Applications to Fetal Programming Studies. *Metabolites* **2022**, *12*, 512. <https://doi.org/10.3390/metabo12060512>

Academic Editor: Karsten Suhre

Received: 29 April 2022

Accepted: 27 May 2022

Published: 2 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The integration of genetics and metabolomics data demands careful accounting of complex dependencies, particularly when modelling familial omics data, e.g., to study fetal programming of related maternal–offspring phenotypes. Efforts to identify genetically determined metabolotypes using classic genome wide association approaches have proven useful for characterizing complex disease, but conclusions are often limited to a series of variant–metabolite associations. We adapt Bayesian network models to integrate metabolotypes with maternal–offspring genetic dependencies and metabolic profile correlations in order to investigate mechanisms underlying maternal–offspring phenotypic associations. Using data from the multiethnic Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study, we demonstrate that the strategic specification of ordered dependencies, pre-filtering of candidate metabolotypes, incorporation of metabolite dependencies, and penalized network estimation methods clarify potential mechanisms for fetal programming of newborn adiposity and metabolic outcomes. The exploration of Bayesian network growth over a range of penalty parameters, coupled with interactive plotting, facilitate the interpretation of network edges. These methods are broadly applicable to integration of diverse omics data for related individuals.

**Keywords:** fetal programming; variant-to-metabolite associations; Bayesian network modeling; serial mediation modeling; integrated multi-omics

## 1. Introduction

Fetal programming describes the series of fetal responses to in utero stimuli, including the maternal genetic and metabolic milieu, during critical phases of cellular differentiation that may substantially impact fetal development [1]. According to the Barker hypothesis, these programmed adaptations may be the origin of a number of diseases in later life [2]. Associations between maternal metabolic phenotypes, including glycemia and body mass index, with newborn adiposity are well documented, e.g., in the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study findings [3,4]. Related investigations using high-dimensional omics data for HAPO mothers and their offspring have provided some clarity around the mechanism of these associations, and include investigations relating maternal genetics and maternal metabolic phenotypes [5], maternal metabolomics and

maternal metabolic phenotypes [6], maternal metabolomics and newborn outcomes [7], fetal genetics and newborn outcomes [8,9], and cord blood metabolomics and newborn outcomes [10]. These analyses have tended to focus on pairs of maternal or newborn predictors and outcomes in HAPO and other cohorts, with only limited attempts at integrating multi-omics data [11].

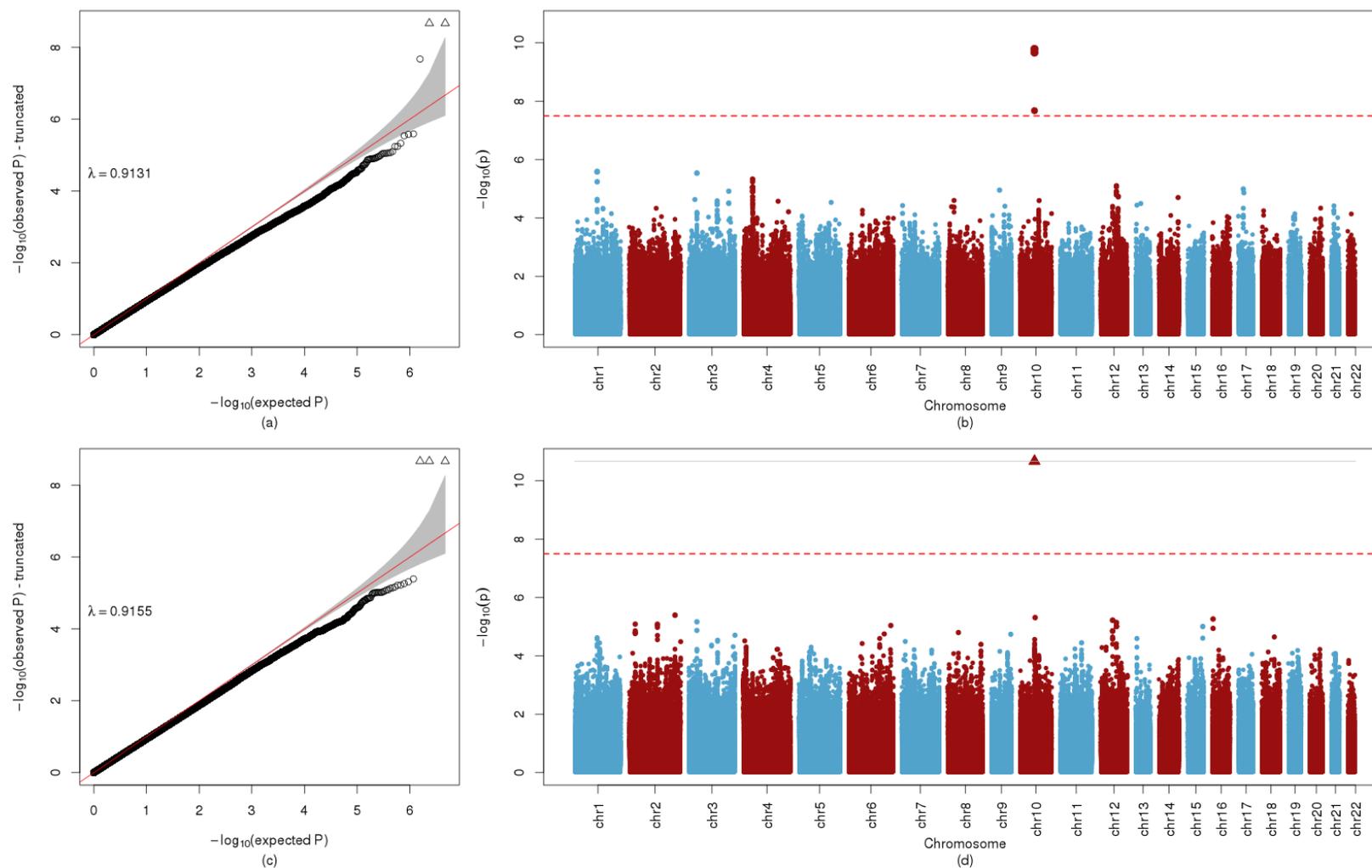
To provide a unified framework for more fully interrogating multi-omics fetal programming contributions that underlie maternal–offspring phenotype associations, we adapt a Bayesian network modeling approach. A Bayesian network is a graphical probabilistic model defined from a set of variables and their conditional dependencies via a directed acyclic graph (DAG). This analytic strategy has been successfully deployed for other analyses of omics data [12–16]. The unique idea behind Bayesian networks is to estimate a series of directed relationships among nodes in which the distribution of a variable represented by a ‘child’ node is described conditional on its ‘parent’ nodes, with a penalty term ‘ $\lambda$ ’ typically applied to induce model sparsity [17]. Ordering of relationships is particularly relevant in multi-omics data analyses. For example, genotypes can affect metabolite levels, but metabolite levels do not affect genotype. In maternal–offspring studies, maternal genotype affects offspring genotype, but not vice versa. Implicit ordering of maternal/newborn metabolite associations is also often identifiable. Directionality of possible associations may be explicitly defined for Bayesian network models, thus accommodating the natural ordering of maternal–offspring and genetic-metabolomics data. Using this cohesive Bayesian network approach, we identify overall trends in the relative contribution of genetics and metabolomics to maternal–offspring phenotypic associations. Additionally, we uncover potential ordered multi-step mediation pathways that are independently validated. This comprehensive multi-omics analytic pipeline for correlated maternal–offspring data provides an alternative to common analytic strategies that emphasize direct associations between omics data sources and phenotypes.

Interpretation and utilization of this scale of integrated omics analysis demands sophisticated graphical display that is accessible to multidisciplinary teams of investigators. To complement Bayesian network estimation, we apply visualization techniques that demonstrate network growth over a range of penalty parameters for the model. In so doing, we observe the varied strengths of multi-omic associations with the HAPO maternal–offspring phenotypes of interest. The approach to comprehensive network visualization is complemented by the ability to highlight specific ordered paths of interest when undertaking strategic analysis of potential mediation mechanisms.

## 2. Results

### 2.1. Significant Variant-to-Metabolite Associations

Across 5.2 million genotyped and imputed SNPs and 331 metabolites (43 unique fasting and 1-hr metabolites from a 2-hr 75g OGTT conducted during HAPO at ~28 weeks’ gestation and/or metabolites from cord blood collected at delivery), a total of 510 unique SNPs had  $p < 10^{-7.5}$  in per-metabolite GWAS analyses. Forty-three pairs of maternal and newborn SNPs remained after LD trimming (Supplemental Table S1). Notable identified associations included *APOA5* (rs2072560 and rs2266788) with fasting maternal triglycerides, *GCKR* with fasting and 1-hr maternal 2-hydroxybutyric acid and maternal 1-hr lactate, and *SLC16A9* (rs1171616, rs1171617, and rs1171619) with fasting and 1-hr maternal AC C2 (Figure 1). The strongest association found was *CPS1* (rs1047891 and another SNP 2500 base-pairs away) with fasting and 1-hr maternal glycine. Three newborn SNPs (rs7601356, rs2286963, and rs3764913) in the *ACADL* region and one newborn SNP (rs3738934) in the *RPE* region were highly associated with cord blood AC C8:1-OH/C6:1-DC. Forty-nine newborn SNPs in the *REV3L* region were highly associated with cord blood NEFA. The 43 pairs of maternal and newborn SNPs listed in Supplemental Table S1 were included as variables for estimating Bayesian network models (BNMs). All significant GWAS results are listed in Supplemental Table S2.



**Figure 1.** QQ and Manhattan Plot AC C2. (a) QQ plot of GWA with fasting AC C2, with corresponding genomic inflation factor,  $\lambda$ . (b) Manhattan plot of GWA with fasting AC C2. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM. (c) QQ plot of GWA with 1-hr AC C2, with corresponding genomic inflation factor,  $\lambda$ . (d) Manhattan plot of GWA with 1-hr AC C2. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM.

## 2.2. Bayesian Network Model (BNM)

Under the highest penalty term value of  $\lambda = 15$  (i.e., the most stringent penalty parameter and thus the sparsest network), the BNM involving fasting maternal glucose and metabolites estimated edges among a network involving 187 nodes with at least one incident edge and 137 edges. As  $\lambda$  values decreased, the number of nodes in each estimated BNM with at least one incident edge and the total number of edges increased, culminating in a BNM involving 309 nodes and 3070 edges for the lowest value of  $\lambda = 1$  (i.e., the least stringent penalty parameter and thus the fullest network). Similarly, for the BNM involving 1-hr metabolites and 1-hr maternal glucose, the most stringent penalty of  $\lambda = 15$  resulted in a BNM with 180 nodes and 132 edges and grew to 306 nodes and 3033 edges at  $\lambda = 1$ . Supplemental Files S1 and S2 demonstrate network growth for networks that were estimating using fasting and 1-hr maternal metabolites, and Table 1 catalogs the numbers of edges connecting nodes of different types at various values of  $\lambda$ .

**Table 1.** Numbers of edges among types of nodes in estimated Bayesian Network Models (BNMs) for a range of  $\lambda$  values.

BNMs Involving Fasting Metabolites and Fasting Maternal Glucose					
Numbers of edges—N (% of total)					
Types of nodes in each edge pair	$\lambda = 15$	$\lambda = 10$	$\lambda = 7$	$\lambda = 2$	$\lambda = 1$
MG-MG	8 (6%)	9 (4%)	11 (3%)	16 (1%)	16 (1%)
MG-MM	0 (0%)	1 (0%)	3 (1%)	293 (10%)	292 (10%)
MG-MP	0 (0%)	0 (0%)	0 (0%)	8 (0%)	22 (1%)
MG-NG	26 (19%)	37 (15%)	38 (9%)	45 (2%)	45 (1%)
MG-NM	0 (0%)	0 (0%)	0 (0%)	114 (4%)	113 (4%)
MG-NP	0 (0%)	0 (0%)	0 (0%)	8 (0%)	8 (0%)
MM-MM	49 (36%)	84 (34%)	159 (37%)	904 (31%)	901 (29%)
MM-MP	0 (0%)	0 (0%)	3 (1%)	8 (0%)	20 (1%)
MM-NM	4 (3%)	18 (7%)	44 (10%)	629 (22%)	627 (20%)
MM-NP	0 (0%)	0 (0%)	0 (0%)	30 (1%)	44 (1%)
MP-MM	0 (0%)	0 (0%)	1 (0%)	31 (1%)	92 (3%)
MP-MP	0 (0%)	0 (0%)	1 (0%)	1 (0%)	1 (0%)
MP-NM	0 (0%)	0 (0%)	0 (0%)	13 (0%)	60 (2%)
MP-NP	0 (0%)	0 (0%)	2 (0%)	4 (0%)	5 (0%)
NG-NG	8 (6%)	9 (4%)	11 (3%)	11 (0%)	11 (0%)
NG-NM	0 (0%)	0 (0%)	0 (0%)	95 (3%)	95 (3%)
NG-NP	0 (0%)	0 (0%)	1 (0%)	6 (0%)	6 (0%)
NM-NM	41 (30%)	90 (36%)	155 (36%)	662 (23%)	661 (22%)
NM-NP	0 (0%)	1 (0%)	1 (0%)	16 (1%)	16 (1%)
NP-NM	0 (0%)	0 (0%)	0 (0%)	30 (1%)	30 (1%)
NP-NP	1 (0%)	1 (0%)	1 (0%)	4 (0%)	5 (0%)

$\lambda = 15$  most stringent penalty,  $\lambda = 1$  least stringent penalty  
 $\lambda = 10$  first occurrence of genotype to non-genotype edge  
 $\lambda = 7$  first occurrence of maternal exposure phenotype to offspring outcome phenotype  
 $\lambda = 2$  first occurrence of SMM pathway evaluated from maternal fasting AC C2 to cord glucose outcome

Table 1. Cont.

BNMs Involving Fasting Metabolites and Fasting Maternal Glucose					
BNMs involving 1-hr Metabolites and 1-hr Maternal Glucose					
Types of nodes in each edge pair	Numbers of edges—N (% of total)				
	$\lambda = 15$	$\lambda = 9$	$\lambda = 7$	$\lambda = 3$	$\lambda = 1$
MG-MG	8 (6.06%)	10 (3.38%)	11 (2.54%)	16 (1.08%)	16 (0.53%)
MG-MM	0 (0%)	1 (0.34%)	3 (0.69%)	89 (6.02%)	272 (8.97%)
MG-MP	0 (0%)	0 (0%)	0 (0%)	8 (0.54%)	34 (1.12%)
MG-NG	26 (19.7%)	37 (12.5%)	38 (8.78%)	45 (3.04%)	45 (1.48%)
MG-NM	0 (0%)	0 (0%)	0 (0%)	21 (1.42%)	93 (3.07%)
MG-NP	0 (0%)	0 (0%)	0 (0%)	1 (0.07%)	10 (0.33%)
MM-MM	43 (32.57%)	105 (35.47%)	156 (36.03%)	507 (34.28%)	893 (29.44%)
MM-MP	0 (0%)	1 (0.33%)	4 (0.92%)	12 (0.81%)	40 (1.32%)
MM-NM	5 (3.79%)	27 (9.12%)	41 (9.47%)	242 (16.36%)	628 (20.71%)
MM-NP	0 (0%)	0 (0%)	1 (0.23%)	14 (0.95%)	47 (1.55%)
MP-MM	0 (0%)	1 (0.34%)	4 (0.93%)	16 (1.08%)	59 (1.95%)
MP-MP	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
MP-NM	0 (0%)	0 (0%)	0 (0%)	4 (0.27%)	55 (1.81%)
MP-NP	0 (0%)	0 (0%)	1 (0.23%)	4 (0.27%)	4 (0.13%)
NG-NG	8 (6.06%)	10 (3.38%)	11 (2.54%)	11 (0.74%)	11 (0.36%)
NG-NM	0 (0%)	0 (0%)	0 (0%)	23 (1.56%)	101 (3.33%)
NG-NP	0 (0%)	0 (0%)	1 (0.23%)	1 (0.07%)	8 (0.26%)
NM-NM	41 (31.06%)	102 (34.46%)	160 (36.95%)	442 (29.88%)	664 (21.89%)
NM-NP	0 (0%)	1 (0.34%)	1 (0.23%)	10 (0.68%)	18 (0.59%)
NP-NM	0 (0%)	0 (0%)	0 (0%)	9 (0.61%)	30 (0.99%)
NP-NP	1 (0.76%)	1 (0.34%)	1 (0.23%)	4 (0.27%)	5 (0.17%)

$\lambda = 15$  most stringent penalty,  $\lambda = 1$  least stringent penalty  
 $\lambda = 9$  first occurrence of genotype to non-genotype edge  
 $\lambda = 7$  first occurrence of maternal exposure phenotype to offspring outcome phenotype  
 $\lambda = 3$  first occurrence of all SMM pathways involving maternal 1-hr metabolites  
MG = maternal genotype  
MM = maternal metabolite  
MP = maternal phenotype  
NG = newborn genotype  
NM = newborn metabolite  
NP = newborn phenotype

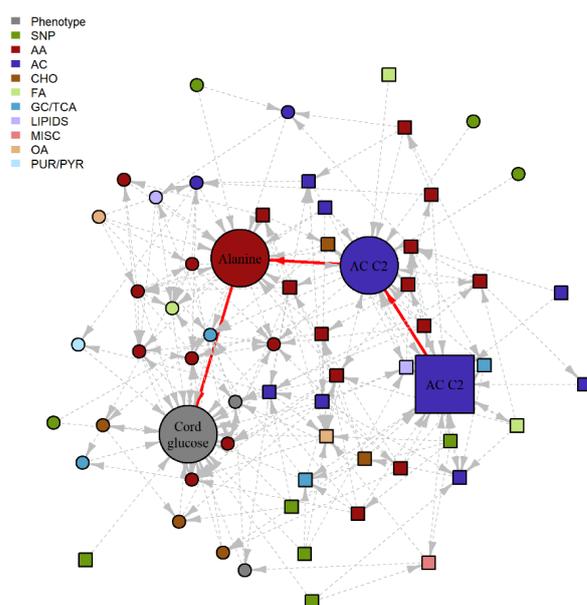
Dynamic animation of both the fasting and 1hr BNMs were generated to better visualize areas of increasing edge density over the range of penalty parameters (Supplemental Files S1 and S2). Nodes were colored to indicate whether they represent phenotypes, genotypes, and metabolites, with different colors for each metabolite compound class. Node shapes were used to indicate whether they represent maternal or offspring variables. As evidenced in Table 1 and in the Supplemental File S1 video animation, most BNM edges exist among maternal metabolites and cord blood metabolites under the most stringent penalty, with edges primarily existing within metabolites of the same class. Connections increase between the maternal and cord blood metabolites before maternal exposure phenotypes and newborn outcome phenotypes connect to the BNM at  $\lambda = 7$  for both the maternal

fasting and 1-hr BNMs. While maternal–offspring genotypes are pairwise connected by edges under the most stringent penalty as would be expected, the first pairwise edge between a genotype node to a non-genotype node appears at  $\lambda = 10$  for the fasting BNM and  $\lambda = 9$  for the 1-hr BNM, where in both instances rs715 from *CPS1* connects to glycine. The *CPS1*-glycine association is well-documented in multiple GWAS studies [18–21]. Substantial connections from genotype nodes to the larger network do not occur until far less stringent  $\lambda$  values are used BNM. This network growth phenomenon suggests that fetal programming mechanisms are likely to be more substantially informed by environmentally driven metabolic contributions than strictly inherited germline genetic variation that may predispose an individual to certain phenotypes.

### 2.3. Serial Mediation Model (SMM)

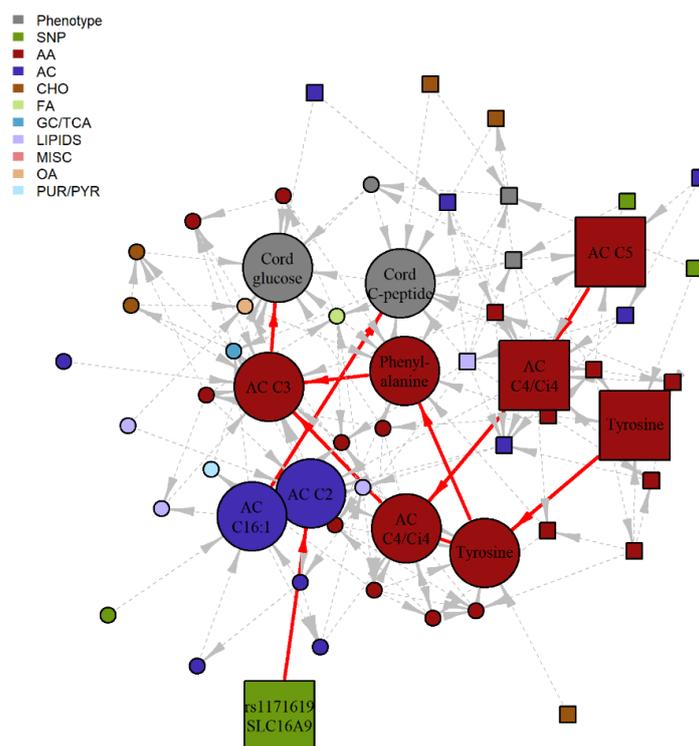
For the fasting and 1-hr BNMs with the least stringent penalty values at  $\lambda = 1$ , all shortest paths from a maternal feature to a newborn outcome (sum of skinfolds (SSF), birthweight, cord C-peptide or cord glucose) were identified. Any path that was a subpath of a larger path was not considered. In total, 3000 unique directed paths from a maternal feature to a newborn outcome were identified in the fasting BNM and 1438 unique paths were identified in the 1-hr BNM. The number of nodes ranged from 3 to 8 in the paths extracted from both the fasting and 1-hr BNMs.

For all identified 4438 different pathways, SMM analyses were performed as described. We specifically examined pathways for which the first node was treated as the independent ‘exposure’ variable and the last node, a newborn outcome (i.e., SSF, birthweight, cord glucose or cord C-peptide), was the dependent ‘outcome’ variable, and all nodes in between were treated as potential mediators, i.e., individual mediators for paths with three nodes or serial mediators for paths with four or more nodes. For the 3000 candidate pathways from the fasting BNM, nine demonstrated a statistically significant IDE after false discovery rate (FDR) adjustment and maintained a positive proportion mediated (PM) in the training data set. For the 1438 candidate pathways from the 1-hr BNM, 22 demonstrated a statistically significant indirect effects (IDE) after FDR adjustment and maintained a positive PM in the training data set. When SMM analyses were repeated in the testing dataset, statistically significant serial mediation was validated for five of these pathways (Figures 2 and 3, Table 2). Notably, only one of the pathways has a genetic component, consistent with previous observations of the stronger metabolomics dependencies in this HAPO data set.



**Figure 2.** Subgraph of fasting BNM highlighting a potential SMM pathway. Nodes represent the metabolite, genotype and phenotype variables and edges represent the conditional dependencies

between them. The maternal features are represented by square nodes and the offspring features are represented by circle nodes. Newborn outcome variables are grey, genotypes are green, and the remaining colors correspond to different metabolite classes. The solid, red edges correspond to the purported SMM pathway with statistically significant IDE after FDR correction and positive proportion mediated in the training data set, with nominal significance and positive PM confirmed in the validation data set. Dashed, grey edges correspond to nodes adjacent to the SMM pathway. AA, amino acid; AC, acylcarnitine; CHO, carbohydrate; FA, fatty acid; GC/TCA, glycolysis/tricarboxylic acid cycle; OA, organic acid; PUR/PYR, purine or pyrimidine.



**Figure 3.** Subgraph of 1-hr BNM highlighting four potential SMM pathway. Nodes represent the metabolite, genotype and phenotype variables and edges represent the conditional dependencies between them. The maternal features are represented by square nodes and the offspring features are represented by circle nodes. Newborn outcome variables are grey, genotypes are green, and the remaining colors correspond to different metabolite classes. The solid, red edges correspond to the four purported SMM pathways involving 1-hr metabolites with statistically significant IDE after FDR correction and positive proportion mediated in the training data set, with nominal significance and positive PM confirmed in the validation data set. Dashed, grey edges correspond to nodes adjacent to the SMM pathways. AA, amino acid; AC, acylcarnitine; CHO, carbohydrate; FA, fatty acid; GC/TCA, glycolysis/tricarboxylic acid cycle; OA, organic acid; PUR/PYR, purine or pyrimidine.

**Table 2.** Individual ordered pathways from Bayesian Network Models (BNMs) demonstrating statistically significant serial mediation in training and validation data sets.

PATHWAY MEMBERS	Training BNM	Validation Data
	IDE (95% Confidence Interval); FDR-Adjusted P, PM (%)	IDE (95% Confidence Interval); Nominal P, PM (%)
Maternal rs1171619 <i>SLC16A9</i> -> Cord blood AC C2 -> Cord blood AC C16:1 -> Cord C-peptide	$-1.60 \times 10^{-2}$ ( $-2.42 \times 10^{-2}$ — $7.78 \times 10^{-3}$ ); 0.0086, 25.51%	$-1.36 \times 10^{-2}$ ( $-2.41 \times 10^{-2}$ — $3.20 \times 10^{-3}$ ); 0.0120, 173.98%

Table 2. Cont.

PATHWAY MEMBERS	Training BNM	Validation Data
	IDE (95% Confidence Interval); FDR-Adjusted P, PM (%)	IDE (95% Confidence Interval); Nominal P, PM (%)
Maternal fasting AC C2 -> Cord blood AC C2 -> Cord blood Alanine -> Cord glucose	$1.26 \times 10^{-2}$ ( $6.12 \times 10^{-3}$ – $1.90 \times 10^{-2}$ ); 0.0202, 129.82%	$1.21 \times 10^{-2}$ ( $1.38 \times 10^{-3}$ – $2.29 \times 10^{-2}$ ); 0.0151, 24.17%
Maternal 1-hr AC C5 -> Maternal 1-hr AC C4/Ci4 -> Cord blood AC C4/Ci4 -> Cord blood AC C3 -> Cord glucose	$-1.21 \times 10^{-2}$ ( $-1.72 \times 10^{-2}$ – $-6.94 \times 10^{-3}$ ); 0.0018, 28.58%	$-8.86 \times 10^{-3}$ ( $-1.45 \times 10^{-2}$ – $-3.23 \times 10^{-3}$ ); 0.0101, 17.63%
Maternal 1-hr Tyrosine -> Cord blood Tyrosine -> Cord blood AC C4/Ci4 -> Cord blood AC C3 -> Cord glucose	$-8.06 \times 10^{-3}$ ( $-1.20 \times 10^{-2}$ – $-4.13 \times 10^{-3}$ ); 0.0053, 55.5%	$-1.13 \times 10^{-2}$ ( $-1.83 \times 10^{-2}$ – $-4.35 \times 10^{-3}$ ); 0.0067, 64.19%
Maternal 1-hr Tyrosine -> Cord blood Tyrosine -> Cord blood Phenylalanine -> Cord blood AC C3 -> Cord glucose	$-7.55 \times 10^{-3}$ ( $-1.16 \times 10^{-2}$ – $-3.48 \times 10^{-3}$ ); 0.0106, 52.02%	$-8.04 \times 10^{-3}$ ( $-1.41 \times 10^{-2}$ – $-2.03 \times 10^{-3}$ ); 0.0261, 45.56%

### 3. Discussion

Integrated modeling of multi-omics data demands careful specification of complex dependencies. For HAPO genetics and metabolomics data, the expected dependencies between observations for mother–offspring pairs merits additional consideration, particularly given the potential for investigating mechanisms underlying fetal programming of related clinical phenotypes in mothers and their newborns using these unique data resources. Recent efforts to identify ‘genetically determined metabolotypes’ using classic GWAS approaches in independent individuals identified SNP-metabolite associations with larger effect sizes than typically observed for clinical phenotypes [19,20,22–24]. Extensions to GWAS, such as phenotype set enrichment analyses to identify SNPs associated with multiple metabolites [25] and informatics resources to investigate purported metabolotypes [26–28], have also proven useful for probing the genetics and metabolomics of complex disease. However, to accurately characterize genetic and metabolic contributions to fetal development, more sophisticated models are required to integrate metabolotypes with maternal–fetal genetic dependencies and metabolic profile correlations. BNMs offer a natural paradigm for this type of data integration due to the ability to naturally incorporate structured, ordered dependencies among different types of data while estimating edges.

A key finding upon global review of our integrated analyses is the relative strength of coordinated behavior among classes of maternal and newborn metabolites compared to all other possible dependencies among the variables included in the estimated BNMs. While we may expect high dependency within maternal and offspring metabolites, the relative density of edges connecting maternal and offspring metabolites under the strictest penalty parameters highlights the tight coordination between the metabolisms of a mother and the developing fetus. Of note, network edges between maternal and offspring metabolites and the maternal and offspring clinical phenotypes emerge prior to any edges between the paired clinical phenotypes (e.g., maternal glucose to cord blood glucose) and shared maternal/offspring genetics. This observation suggests that environmentally derived or influenced metabolomic contributions may outweigh the genetic underpinnings of fetal programming for size and adiposity at birth.

Complementing the big picture linking of metabolomics and genetics contributions to maternal/offspring phenotypes is the specific pathway-level interpretation of possible mediating mechanisms. Serial mediation analyses informed by comprehensive BNM estimation indicated biologically plausible mediation mechanisms. In the pathway from 1-hr tyrosine to cord glucose via cord blood tyrosine, phenylalanine and propionyl carnitine (C3), the BNM suggests that higher levels of maternal 1-hr tyrosine indirectly result in lower cord glucose and revealed the plausible intermediate pathways linking these two. It is known that phenylalanine is an essential amino acid and is metabolized to tyrosine when the tyrosine level is low. However, if maternal tyrosine is high and is transported across the placenta leading to sufficient fetal tyrosine, then the conversion of phenylalanine to tyrosine is reduced. C3 is formed from amino acid metabolism but primarily branch chain amino acids (BCAA). Previous studies have shown that the aromatic acids, phenylalanine, tyrosine, and the BCAA, are associated with insulin resistance in adults, including in pregnancy [11]. Thus, it is plausible that they could be involved in fetal glycemic regulation. For the pathway from *SLC16A9* to cord C-peptide via cord blood acetylcarnitine (C2) and hexadecenoylcarnitine (C16:1), *SLC16A9* is a transporter of monocarboxylates that has been associated with C2 levels [29]. Elevated C2 levels in the mother may affect C2 in the fetus through placental transfer. While the direct relationship between C2 and C16:1 is unclear, both C2 and C16:1 levels are higher in newborns with extreme macrosomia [24] and higher levels of C16:1 have been noted in newborns of mothers with GDM [30]. Notably, the biological plausibility of these purported SMM pathways is based on observations in adult metabolism. Our findings when jointly examining maternal and newborn data with the described dependencies suggest similar pathway involvement very early in life. Furthermore, examination of serial pathways in the BNM highlights that while maternal glucose is typically emphasized when studying newborn adiposity and glucose/insulin regulation, in fact multiple maternal metabolic factors, including metabolites that may cross the placenta, are part of the fuller picture.

Penalty parameter selection for sparse estimation of edges in large scale networks is a topic of focused discussion and ongoing bioinformatics research. In contrast to selecting a single penalty parameter that optimizes specific criteria for estimation of one BNM of our integrated data, we opted to view and animate BNMs over a range of penalty parameters. This approach to visualization, along with color distinctions for node types, naturally facilitated the observation of the relative centrality of maternal/offspring metabolite dependencies compared to shared genetics for explaining dependencies between clinical phenotypes that were specifically investigated in HAPO. For other studies involving multi-omics data, exploring network growth over decreasing stringency requirements may illuminate the relative strengths and cooperativity of different types of omics features.

There are limitations to our analytic approach. In our BNM analyses, we blocklisted potential edges from offspring SNPs to maternal phenotypes and fasting/1-hr maternal metabolites such that no such edge could be included in the estimated models. However, there is evidence that offspring genotype can have limited influence maternal phenotypes [31,32]. Implementing the specified blocklist prevented the discovery of such rare occurrences with our data. We chose this condition to reduce noise from these edges for BNM estimation. A second limitation pertains to the examination of serial mediation in the BNM as independent paths. As shown, there are two purported mediation pathways from maternal 1-hr tyrosine to cord glucose. Another approach on this specific pathway would be to use a combination of both parallel and serial mediation modeling. In addition, we focused only on serial mediation paths that started at a maternal variable and ended at a newborn outcome. Other routes of investigation for different combinations of pathway starting and end points may yield further insight about mediating mechanisms.

In summary, Bayesian network modeling and animated visualization over a range of penalty parameters provided a natural analytic framework for the cogent synthesis of genetics and metabolomics data related to clinical phenotypes for HAPO mothers and their newborns. This approach facilitated both large-scale summaries of the relative contribu-

tions of data types, as well as local interrogation of potential serial mediating pathways. These pathways may suggest unique lines for further laboratory and/or population-level investigation, potentially incorporating other edges from the BNM that influence the linear serial mediating pathway of primary interest. Beyond the HAPO study, the approaches described here can be adapted to integrate omics data for other studies that are designed to evaluate complex contributions that may underlie clinical phenotypic associations.

## 4. Materials and Methods

### 4.1. Data Sets

#### 4.1.1. Hyperglycemia and Adverse Pregnancy Outcome Study

The HAPO Study was an observational, population-based study conducted from 2000–2006 at 15 international field centers that recruited over 25,000 participants [3]. Enrolled women underwent a 75-g oral glucose tolerance test (OGTT) at approximately 28 weeks' gestation after an overnight fast. Fasting, 1-hr, and 2-hr glucose levels were measured as described [33,34]. Maternal blood samples during the OGTT were placed on ice immediately, processed within 60 min of collection, stored at  $-20\text{ }^{\circ}\text{C}$  or  $-80\text{ }^{\circ}\text{C}$  for 1–6 weeks, shipped on dry ice to the HAPO Central Laboratory, and stored at  $-80\text{ }^{\circ}\text{C}$ . Each participant's age, gestational age at time of OGTT, family history of diabetes and hypertension, parity, and history of cigarette smoking and alcohol use were obtained by questionnaire. Cord blood samples were obtained within 5 min of delivery and before delivery of the placenta, processed within 30 min, and stored at  $-70\text{ }^{\circ}\text{C}$  until laboratory assays were performed. Cord blood C-peptide, birthweight, and sum of skinfolds (SSF) were measured as described using calibrated equipment and standardized methods [33–35]. The HAPO protocol was approved by each field center's institutional review board, and participants gave informed consent.

#### 4.1.2. Metabolomics Data

Conventional clinical and targeted metabolites were measured in 3463 maternal fasting and 1-hr serum samples [36] and in 1600 offspring cord blood serum samples [10,37]. Conventional clinical metabolites (triglycerides, nonesterified fatty acid (NEFA), lactate, glycerol, and 3-hydroxybutyrate) were measured using standard enzymatic chemistries on a Beckman-Coulter Unicell DxC 600 clinical analyzer. Targeted panels of amino acids and acylcarnitines were analyzed by flow-injection, electrospray-ionization tandem mass spectrometry (MS) and quantified by isotope or pseudo-isotope dilution using a Waters TQ triple quadrupole MS, equipped with an Acquity liquid chromatography system, and with data handling in the MassLynx 4.1 environment (Waters Corporation, Milford, MA, USA). Five conventional clinical metabolites, 15 amino acids, and 45 acylcarnitines were analyzed across maternal fasting, 1-hr, and offspring cord blood. Targeted metabolites with >10% missing data across all samples were excluded from analysis, yielding a total of 57, 56, and 48 targeted metabolites for analysis in maternal fasting, 1-hr, and offspring cord blood, respectively.

Non-targeted assays were performed using gas chromatography-mass spectrometry (GC-MS) as described [38]. Briefly, maternal fasting, 1-hr, and offspring cord blood serum samples were extracted with methanol that was spiked with a retention time-locking internal standard of perdeuterated myristic acid. After methanol extraction, drying and derivatization by methoximation and trimethylsilylation, samples were run in daily batches of matched sets of fasting and 1-hr maternal OGTT sera and cord blood sera on a 7890B GC/5977B MS (Agilent Technologies, Santa Clara, CA, USA). Each batch also included multiple injections of a quality control (QC) sample and a process blank. QC samples consisted of uniform pools generated from aliquots of all sera in the study and were injected at the beginning, middle and end of each batch. GC-MS peaks were deconvoluted with AMDIS freeware [39] (<http://www.amdis.net/> accessed on 22 June 2021 [40]) and annotated using the Agilent Fiehn GC-MS Metabolomics RTL Library [41] with additions from the laboratory at Duke University School of Medicine. Only annotated peaks were included

for analysis. Integrated peak areas were  $\log_2$  transformed. QC data were then used to account for technical variability on a feature-specific basis, with batch correction and normalization performed as described using the *metabomxtr* R package (version 1.26.0) [42,43]. Non-targeted metabolites with >20% missing data across all samples were excluded from analysis. Remaining non-targeted metabolites with <20% missing data were imputed using metabolite-specific minimum values since, in general, non-targeted metabolite missing values are attributable to low abundance. Non-targeted metabolite measures that were also represented in the targeted assays were excluded from analysis. In total, 55, 53, and 47 non-targeted metabolites were included in the final analyses from the maternal fasting, 1-hr, and offspring cord blood samples, respectively.

#### 4.1.3. Genotype Data

Genotyping and QC for HAPO data have been described for an initial set of 2581 maternal and newborn Afro-Caribbean and 1615 maternal and newborn Mexican American DNA samples (Illumina HumanOmni1M-Duo v3B SNP array, Illumina, San Diego, CA, USA), 3152 maternal and newborn Northern European samples (Illumina Human610-Quad v1B SNP array, run at the Broad Institute, Cambridge, MA, USA), and 2466 maternal and newborn Thai samples (Illumina HumanOmni-Quad v1-0B SNP array, run at the Center for Inherited Disease Research, Baltimore, MD, USA) [5,9].

In addition, a subset of 2680 Northern European maternal and newborn samples were run using the Illumina Human 610 Quad v1 B SNP array at the Broad Institute Center for Genotyping and Analysis (CGA), USA [44], and a total of 5564 maternal and newborn transethnic samples were processed on the Illumina Global Screening Array-24 v2.0 A1 following agreed-upon protocols of the Gene, Environment Associations Studies consortium [45]. Exclusion criteria for samples and SNPs included inconsistent self-report of sex at birth with available chromosomal data, chromosomal anomalies, unintended sample duplicates, sample relatedness based on genetic relatedness, low call rate (samples with missing call rate >1% and SNPs with missing call rate >2%), SNPs with >3 Mendelian errors, departures from Hardy–Weinberg equilibrium ( $< 1 \times 10^{-4}$  for the original Afro-Caribbean, Mexican American, Northern European, and Thai groups,  $< 1 \times 10^{-6}$  for the additional 2680 Northern European group, and  $< 1 \times 10^{-3}$  for the transethnic group), duplicate discordance, sex differences in heterozygosity, low minor allele frequencies (<0.01) and/or low imputation quality score (<0.75).

Genotype imputation was performed separately on the twelve datasets (six maternal and six fetal) on the TOPMed Imputation Server using Minimac4 (version 1.5.7) [46] and the TOPMed reference panel for the Afro-Caribbean, Mexican American, Northern European, Thai, Northern European and transethnic samples [47]. Consistent strand assignments between the reference dataset and the QC-cleaned and filtered datasets were ensured using the strand-checking utility of McCarthy Group Tools. Strand was corrected and/or SNPs were removed where strandedness could not be resolved. The HAPO genotype was then imputed to the TOPMed reference panel, using an  $r^2$  filter of 0.3 to remove unreliably imputed SNPs. A total of 5,237,318 overlapping SNPs were analyzed in a mega-analysis across the twelve imputed datasets. Principle components (PCs) were estimated separately for maternal and offspring data using the package *SNPRelate* version 1.26.0 in R [48].

#### 4.1.4. Combined Genotype, Metabolomics and Phenotype Training and Validation Data Sets

For analyses reported here, 1385 HAPO mother/offspring pairs with all of the following data types were retained for analysis: (1) maternal genotype, (2) offspring genotype, (3) maternal fasting and 1-hr targeted and non-targeted metabolomics, (4) offspring cord blood targeted and non-targeted metabolomics, (5) maternal fasting and 1-hr glucose measurements and maternal BMI from the HAPO OGTT at ~28 weeks' gestation along with additional maternal data used for model adjustments, and (6) offspring sum of skinfolds, birthweight, cord glucose, and cord C-peptide measured at delivery. Two-thirds of these

samples ( $n = 924$ ) were used as a training data set for Bayesian network model development and selection of potential mediation pathways. The remaining one-third ( $n = 461$ ) were used as a validation data set for confirming statistical significance of purported mediation findings. Sample selection for training and validation data sets was stratified by ancestry group based on genotype data.

## 4.2. Statistical Analyses

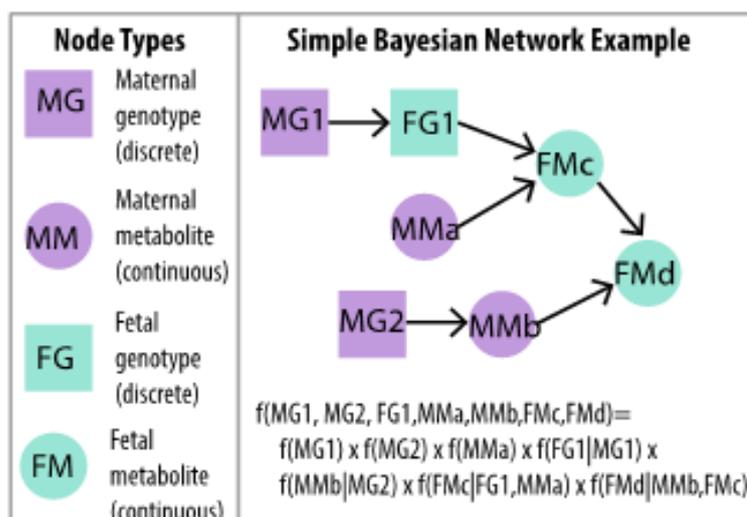
### 4.2.1. Variant-to-Metabolite Associations

In order to reduce computational burden when estimating Bayesian networks, an initial round of ‘metabotyping’ was performed, i.e., genome wide association (GWAS) analyses were performed separately for all metabolites as outcomes [22,26]. Metabotyping was performed on the combined training and validation data sets using a mega-analysis linear regression approach with SNPTest v2.5 against an additive genotype variable, separately for maternal fasting and 1-hr metabolites with maternal genotype, and offspring cord blood metabolites with offspring genotype. Maternal fasting metabolite models were adjusted for the first 3 principal components of genetic ancestry, maternal age, gestational age, BMI, height, parity (0/1+), mean arterial pressure and fasting plasma glucose (1-hr glucose for 1-hr metabolites) at HAPO OGTT, and sample storage time. Newborn cord blood metabolites were adjusted for the first 3 principal components of genetic ancestry, maternal age, BMI, height, parity (0/1+), mean arterial pressure at OGTT, gestational age at delivery, newborn sex, and sample storage time. SNPs with nominal  $p < 10^{-7.5}$  were selected for further analysis. This selection threshold was set to be slightly more inclusive than typical genome-wide significance of nominal  $p < 10^{-8}$ , anticipating that the penalized network modeling approach would aid in the selection the most informative SNPs. Each metabolite with significant SNP associations was LD trimmed independently using the package *SNPRelate* version 1.26.0 in R [48].

### 4.2.2. Bayesian Network Modeling (BNM)

We estimated BNMs including the following variables as nodes in the network: maternal exposure variables, namely BMI at OGTT,  $\log_{10}$  fasting plasma glucose at OGTT, square root of 1-hr plasma glucose at OGTT (fasting and 1-hr glucose transformations were consistent with previous GWAS analyses of maternal phenotypes [5]), the 43 SNPs identified in preliminary metabotyping, 117 fasting maternal metabolites, 114 1-hr maternal metabolites, 100 cord blood metabolites, and a set of newborn anthropometric outcomes (birthweight, square root sum of skinfold,  $\log_{10}$  cord glucose, and  $\log_{10}$  cord C-peptide; data transformations were consistent with previous GWAS analyses of these newborn phenotypes [9]). A simple schematic illustrating the types of ordered relationships that might be estimated among maternal and fetal genotypes and metabolites and a corresponding set of conditional dependencies that are central to BNM estimation are illustrated in Figure 4. BNMs were estimated using the training data set, as specified above. We used a fast, score-based method that uses sparse regularization and block-cyclic coordinate descent approach to build these BNMs [17]. Conditions on edge directionality were specified using a ‘blocklist’, with the following edge types omitted from the range of potential estimated edges: offspring SNPs to maternal SNPs; newborn outcomes to maternal SNPs; cord blood metabolites to maternal SNPs; offspring SNPs to maternal phenotypes; offspring SNPs to fasting/1-hr maternal metabolites; newborn outcomes to offspring SNP; cord blood metabolites to offspring SNPs; maternal phenotypes to offspring SNPs; fasting/1-hr maternal metabolites to maternal SNPs; fasting/1-hr maternal metabolites to offspring SNPs; and SNPs from one chromosome to SNPs in different chromosomes. A vector of  $\lambda$  values was used in the solution path, starting at  $\lambda = 15$  and decreasing to  $\lambda = 1$ , and networks were estimated across a range of these regularization penalty parameters. Separate BNMs were estimated for fasting maternal metabolites/fasting maternal glucose and for their 1-hr counterparts. Since metabolism at the fasting and 1-hr post-glucose timepoints may reflect

different components of glucose metabolism, we chose to model data from the fasting and 1-hr timepoints separately.



**Figure 4.** Simple BNM example of two maternal genotypes (MG1, MG2), one fetal genotype at the same locus as one maternal genotype (FG1), two maternal metabolites (MMa, MMb) and two other fetal metabolites (FMc, FMd) in which the full multivariate distribution is written as a product of conditional distributions. Ordered dependencies among maternal and fetal genotypes and metabolites are naturally accommodated in BNM estimation.

#### 4.2.3. Serial Mediation Model (SMM)

Shortest directed paths in the estimated Bayesian network under the lowest  $\lambda$  value (i.e., least stringent penalty parameter) were identified as possible candidates for mediation analyses using shortest path identification functionality in the *igraph* R package [49]. In particular, we focused on shortest directed paths connecting a maternal SNP, fasting/1-hr metabolite, or maternal phenotype as a starting point along a directed path to a newborn outcome as the end point of the path. These paths were interpreted as purported mediation pathways connecting a maternal exposure to a newborn phenotype and thus potentially informative for investigating mechanisms of fetal programming. All paths containing 3 or more nodes were examined using serial mediation model (SMM) analysis with structural equation modeling using the *lavaan* R package version 0.6-11 [50]. The serial mediator Model 6 from Model Templates for PROCESS for SPSS and SAS [51] was adapted to test statistical significance of serial mediation through different pathways involving maternal genetics, maternal fasting/1-hr metabolomics, maternal phenotypes, offspring genetics, and cord blood metabolomics to inform a newborn outcome. Indirect effects (IDE), direct effects (DE), and proportion mediated (PM) were estimated and standard errors were estimated through bootstrapping. False discovery rate adjustment was applied to IDE  $p$ -values separately for the fasting BNM and 1-hr BNM paths [52]. SMMs were adjusted for the following covariates at the HAPO OGTT: field center, maternal age, gestational age, height, parity (0/1+), mean arterial pressure, sample storage time, offspring sex, and gestational age at delivery. SMM analysis was on performed on the training dataset. Those pathways with significant indirect effects Benjamini-Hochberg adjusted  $p$ -values  $< 0.05$  and a positive proportion mediated were evaluated for replication in the testing dataset according to nominal  $p$ -value  $< 0.05$ .

#### 4.2.4. Network Visualization

We generated interactive visualizations to assist in the interpretation of BNMs over a range of penalty parameters using functions from the *igraph* version 1.2.4.1 [49], *intergraph* version 2.0-2 [53], *networkDynamic* version 0.10.0 [54], *network* version 1.15 [55,56], and *ndtv* version 0.13.0 [57] R packages. Graphs were animated over 15 penalty terms starting at

$\lambda = 15$  and decreasing to  $\lambda = 1$  to demonstrate network growth as the penalty decreased. The  $x$ -axis on the.html files that contain the network animations index the increasing density of the BNM estimates, i.e., an index of 1 on the  $x$ -axis in the animation corresponds to  $\lambda = 15$ , an index of 2 to  $\lambda = 14$ , and so on.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/metabo12060512/s1>, Figure S1: QQ and Manhattan Plot glycine (a) QQ plot of GWA with fasting glycine, with corresponding genomic inflation factor,  $\lambda$ . (b) Manhattan plot of GWA with fasting glycine. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM. (c) QQ plot of GWA with 1-hr glycine, with corresponding genomic inflation factor,  $\lambda$ . (d) Manhattan plot of GWA with 1-hr glycine. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM; Figure S2: QQ and Manhattan Plot 2-hydroxybutyric acid (a) QQ plot of GWA with fasting 2-hydroxybutyric acid, with corresponding genomic inflation factor,  $\lambda$ . (b) Manhattan plot of GWA with fasting 2-hydroxybutyric acid. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM. (c) QQ plot of GWA with 1-hr 2-hydroxybutyric acid, with corresponding genomic inflation factor,  $\lambda$ . (d) Manhattan plot of GWA with 1-hr 2-hydroxybutyric acid. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM; Figure S3: QQ and Manhattan Plot triglycerides and lactate (a) QQ plot of GWA with fasting triglycerides, with corresponding genomic inflation factor,  $\lambda$ . (b) Manhattan plot of GWA with fasting triglycerides. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM. (c) QQ plot of GWA with 1-hr lactate, with corresponding genomic inflation factor,  $\lambda$ . (d) Manhattan plot of GWA with 1-hr lactate. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM; Figure S4: QQ and Manhattan Plot AC C5-OH/C3-DC and NEFA (a) QQ plot of GWA with cord blood AC C5-OH/C3-DC, with corresponding genomic inflation factor,  $\lambda$ . (b) Manhattan plot of GWA with cord blood AC C5-OH/C3-DC. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM. (c) QQ plot of GWA with cord blood NEFA, with corresponding genomic inflation factor,  $\lambda$ . (d) Manhattan plot of GWA with cord blood NEFA. The red line indicates  $-\log_{10}p = 7.5$ , the criterion used for potential inclusion in the BNM; Table S1: Forty-three maternal and newborn SNPs (GRCh38); Table S2: Significant variant-to-metabolite associations; File S1: Maternal Fasting Metabolites BNM; File S2: Maternal 1-hr Metabolites BNM.

**Author Contributions:** Conceptualization, A.K. and D.M.S.; methodology, A.K. and D.M.S.; software, A.K.; formal analysis, A.K.; data curation, A.K.; writing—original draft preparation, A.K. and D.M.S.; writing—review and editing, A.K., D.M.S., M.-F.H., M.G.H., R.B. and W.L.L.J.; visualization, A.K.; supervision, D.M.S.; funding acquisition, D.M.S., R.B. and W.L.L.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** The HAPO Study was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant Nos. R01HD34242 and R01HD34243), with additional HAPO ancillary study data obtained through the National Institute of Diabetes and Digestive and Kidney Diseases (grant Nos. R01DK095963 and R01DK117491). The metabolotyping and network analyses described in this work were funded by the National Library of Medicine (grant No. R01LM013444) and the National Cancer Institute (grant No. R03CA211318).

**Institutional Review Board Statement:** The HAPO study protocol was approved by the institutional review board at each field center (protocol code 353-003 and date of approval 28 June 2000).

**Informed Consent Statement:** All participants gave written informed consent.

**Data Availability Statement:** The maternal and newborn genotype data on European ancestry, Mexican American, Thai and Afro-Caribbean proposed for use in these studies and the accompanying phenotype data are currently available through dbGaP ([www.ncbi.nlm.nih.gov/gap](http://www.ncbi.nlm.nih.gov/gap), accessed on 1 June 2021). The remaining additional subset of 2680 Northern European and transethnic samples. metabolomic data and codes used for analyses will be made available by the authors upon request.

**Acknowledgments:** This paper is published on behalf of the HAPO Study Cooperative Research Group.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kwon, E.J.; Kim, Y.J. What is fetal programming? A lifetime health is under the control of in utero health. *Obstet. Gynecol. Sci.* **2017**, *60*, 506–519. [[CrossRef](#)] [[PubMed](#)]
2. Barker, D.J. In utero programming of chronic disease. *Clin. Sci.* **1998**, *95*, 115–128. [[CrossRef](#)]
3. HAPO Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcomes. *N. Engl. J. Med.* **2008**, *358*, 1991–2002. [[CrossRef](#)] [[PubMed](#)]
4. HAPO Study Cooperative Research Group. Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study: Associations with maternal body mass index. *BJOG* **2010**, *117*, 575–584. [[CrossRef](#)] [[PubMed](#)]
5. Hayes, M.G.; Urbanek, M.; Hivert, M.F.; Armstrong, L.L.; Morrison, J.; Guo, C.; Lowe, L.P.; Scheftner, D.A.; Pluzhnikov, A.; Levine, D.M.; et al. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes* **2013**, *62*, 3282–3291. [[CrossRef](#)] [[PubMed](#)]
6. Liu, Y.; Kuang, A.; Bain, J.R.; Muehlbauer, M.J.; Ilkayeva, O.R.; Lowe, L.P.; Metzger, B.E.; Newgard, C.B.; Scholtens, D.M.; Lowe, W.L. Maternal Metabolites Associated With Gestational Diabetes Mellitus and a Postpartum Disorder of Glucose Metabolism. *J. Clin. Endocrinol. Metab.* **2021**, *106*, 3283–3294. [[CrossRef](#)]
7. Kadakia, R.; Nodzinski, M.; Talbot, O.; Kuang, A.; Bain, J.R.; Muehlbauer, M.J.; Stevens, R.D.; Ilkayeva, O.R.; O’Neal, S.K.; Lowe, L.P.; et al. Maternal metabolites during pregnancy are associated with newborn outcomes and hyperinsulinaemia across ancestries. *Diabetologia* **2019**, *62*, 473–484. [[CrossRef](#)]
8. Hughes, A.E.; Nodzinski, M.; Beaumont, R.N.; Talbot, O.; Shields, B.M.; Scholtens, D.M.; Knight, B.A.; Lowe, W.L., Jr.; Hattersley, A.T.; Freathy, R.M. Fetal Genotype and Maternal Glucose Have Independent and Additive Effects on Birth Weight. *Diabetes* **2018**, *67*, 1024–1029. [[CrossRef](#)]
9. Urbanek, M.; Hayes, M.G.; Armstrong, L.L.; Morrison, J.; Lowe, L.P.; Badon, S.E.; Scheftner, D.A.; Pluzhnikov, A.; Levine, D.; Laurie, C.C.; et al. The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum. Mol. Genet.* **2013**, *22*, 3583–3596. [[CrossRef](#)]
10. Kadakia, R.; Scholtens, D.M.; Rouleau, G.W.; Talbot, O.; Ilkayeva, O.R.; George, T.; Josefsen, J.L. Cord Blood Metabolites Associated with Newborn Adiposity and Hyperinsulinemia. *J. Pediatr.* **2018**, *203*, 144–149.e141. [[CrossRef](#)]
11. Liu, Y.; Kuang, A.; Talbot, O.; Bain, J.R.; Muehlbauer, M.J.; Hayes, M.G.; Ilkayeva, O.R.; Lowe, L.P.; Metzger, B.E.; Newgard, C.B.; et al. Metabolomic and genetic associations with insulin resistance in pregnancy. *Diabetologia* **2020**, *63*, 1783–1795. [[CrossRef](#)] [[PubMed](#)]
12. Friedman, N.; Linial, M.; Nachman, I.; Pe’er, D. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **2000**, *7*, 601–620. [[CrossRef](#)] [[PubMed](#)]
13. Friedman, N.; Nachman, I.; Pe’er, D. Learning bayesian network structure from massive datasets: The Sparse Candidate algorithm. *Proc. Fifteenth Conf. Uncertain. Artif. Intell.* **1999**, 206–215. [[CrossRef](#)]
14. Koch, M.; Broom, B.M.; Subramanian, D. Learning robust cell signalling models from high throughput proteomic data. *Int. J. Bioinform. Res. Appl.* **2009**, *5*, 241–253. [[CrossRef](#)] [[PubMed](#)]
15. Ruiz-Perez, D.; Lugo-Martinez, J.; Bourguignon, N.; Mathee, K.; Lerner, B.; Bar-Joseph, Z.; Narasimhan, G. Dynamic Bayesian Networks for Integrating Multi-omics Time Series Microbiome Data. *mSystems* **2021**, *6*. [[CrossRef](#)] [[PubMed](#)]
16. Wang, Q.; Chen, R.; Cheng, F.; Wei, Q.; Ji, Y.; Yang, H.; Zhong, X.; Tao, R.; Wen, Z.; Sutcliffe, J.S.; et al. A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nat. Neurosci.* **2019**, *22*, 691–699. [[CrossRef](#)]
17. Aragam, B.; Zhou, Q. Concave Penalized Estimation of Sparse Gaussian Bayesian Networks. *J. Mach. Learn. Res.* **2015**, *16*, 2273–2328.
18. Hartiala, J.A.; Wilson Tang, W.H.; Wang, Z.; Crow, A.L.; Stewart, A.F.R.; Roberts, R.; McPherson, R.; Erdmann, J.; Willenborg, C.; Hazen, S.L.; et al. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat. Commun.* **2016**, *7*, 10558. [[CrossRef](#)]
19. Illig, T.; Gieger, C.; Zhai, G.; Romisch-Margl, W.; Wang-Sattler, R.; Prehn, C.; Altmaier, E.; Kastenmuller, G.; Kato, B.S.; Mewes, H.-W.; et al. A genome-wide perspective of genetic variation in human metabolism. *Nat. Genet.* **2010**, *42*, 137–141. [[CrossRef](#)]
20. Suhre, K.; Shin, S.-Y.; Petersen, A.-K.; Mohny, R.P.; Meredith, D.; Wagele, B.; Altmaier, E.; CARDIoGRAM; Deloukas, P.; Erdmann, J.; et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **2011**, *477*, 54–60. [[CrossRef](#)]
21. Mittelstrass, K.; Ried, J.S.; Yu, Z.; Krumsiek, J.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Polonikov, A.; Peters, A.; Theis, F.J.; et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **2011**, *7*, e1002215. [[CrossRef](#)] [[PubMed](#)]
22. Gieger, C.; Geistlinger, L.; Altmaier, E.; Hrabce de Angelis, M.; Kronenberg, F.; Meitinger, T.; Mewes, H.-W.; Wichmann, H.-E.; Weinberger, K.M.; Adamski, J.; et al. Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **2008**, *4*, e1000282. [[CrossRef](#)] [[PubMed](#)]
23. Jourdan, C.; Petersen, A.-K.; Gieger, C.; Doring, A.; Illig, T.; Wang-Sattler, R.; Meisinger, C.; Peters, A.; Adamski, J.; Prehn, C.; et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS ONE* **2012**, *7*, e40009. [[CrossRef](#)] [[PubMed](#)]
24. Raffler, J.; Friedrich, N.; Arnold, M.; Kacprowski, T.; Rueedi, R.; Altmaier, E.; Bergmann, S.; Budde, K.; Gieger, C.; Homuth, G.; et al. Genome-wide association study with targeted and non-targeted NMR metabolomics identifies 15 novel loci of urinary human metabolic individuality. *PLoS Genet.* **2015**, *11*, e1005487. [[CrossRef](#)]

25. Ried, J.S.; Shin, S.-Y.; Krumsiek, J.; Illig, T.; Theis, F.J.; Spector, T.D.; Adamski, J.; Wichmann, H.-E.; Strauch, K.; Soranzo, N.; et al. Novel genetic associations with serum level metabolites identified by phenotype set enrichment analyses. *Hum. Mol. Genet.* **2014**, *23*, 5847–5857. [[CrossRef](#)]
26. Shin, S.-Y.; Fauman, E.B.; Petersen, A.-K.; Krumsiek, J.; Santos, R.; Huang, J.; Arnold, M.; Erte, I.; Forgetta, V.; Yang, T.-P.; et al. An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **2014**, *46*, 543–550. [[CrossRef](#)]
27. Arnold, M.; Raffler, J.; Pfeufer, A.; Suhre, K.; Kastenmuller, G. SNIIPA: An interactive, genetic variant-centered annotation browser. *Bioinformatics* **2015**, *31*, 1334–1336. [[CrossRef](#)]
28. Kastenmuller, G.; Raffler, J.; Gieger, C.; Suhre, K. Genetics of human metabolism: An update. *Hum. Mol. Genet.* **2015**, *24*, R93–R101. [[CrossRef](#)]
29. Feofanova, E.V.; Chen, H.; Dai, Y.; Jia, P.; Grove, M.L.; Morrison, A.C.; Qi, Q.; Daviglus, M.; Cai, J.; North, K.E.; et al. A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos. *Am. J. Hum. Genet.* **2020**, *107*, 849–863. [[CrossRef](#)]
30. Herrera-Van Oostdam, A.S.; Salgado-Bustamante, M.; Lima-Rogel, V.; Oropeza-Valdez, J.J.; López, J.A.; Rodríguez, I.D.R.; Toro-Ortiz, J.C.; Herrera-Van Oostdam, D.A.; López-Hernández, Y.; Monárrez-Espino, J. Urinary Metabolomic Profile of Neonates Born to Women with Gestational Diabetes Mellitus. *Metabolites* **2021**, *11*, 723. [[CrossRef](#)]
31. Procopciuc, L.M.; Caracostea, G.; Zaharie, G.; Stamatian, F. Newborn APOE genotype influences maternal lipid profile and the severity of high-risk pregnancy-preeclampsia: Interaction with maternal genotypes as a modulating risk factor in preeclampsia. *Hypertens Pregnancy* **2015**, *34*, 271–283. [[CrossRef](#)]
32. Maria Procopciuc, L.; Caracostea, G.; Maria Hazi, G.; Nemeti, G.; Zaharie, G.; Stamatian, F. Maternal/fetal eNOS-Glu298Asp genotypes and their influence on the severity, prognosis, and lipid profile of preeclampsia. *J. Matern.-Fetal Neonatal Med.* **2018**, *31*, 1681–1688. [[CrossRef](#)] [[PubMed](#)]
33. Nesbitt, G.S.; Smye, M.; Sheridan, B.; Lappin, T.R.J.; Trimble, E.R.; HAPO Study Cooperative Research Group. Integration of local and central laboratory functions in a worldwide multicentre study: Experience from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. *Clin. Trials* **2006**, *3*, 397–407. [[PubMed](#)]
34. HAPO Study Cooperative Research Group. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. *Int. J. Gynaecol. Obstet.* **2002**, *78*, 69–77. [[CrossRef](#)]
35. HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: Associations with neonatal anthropometrics. *Diabetes* **2009**, *58*, 453–459. [[CrossRef](#)] [[PubMed](#)]
36. Jacob, S.; Nodzinski, M.; Reisetter, A.C.; Bain, J.R.; Muehlbauer, M.J.; Stevens, R.D.; Ilkayeva, O.R.; Lowe, L.P.; Metzger, B.E.; Newgard, C.B.; et al. Targeted Metabolomics Demonstrates Distinct and Overlapping Maternal Metabolites Associated With BMI, Glucose, and Insulin Sensitivity During Pregnancy Across Four Ancestry Groups. *Diabetes Care* **2017**, *40*, 911–919. [[CrossRef](#)] [[PubMed](#)]
37. Kadakia, R.; Talbot, O.; Kuang, A.; Bain, J.R.; Muehlbauer, M.J.; Stevens, R.D.; Ilkayeva, O.R.; Lowe, L.P.; Metzger, B.E.; Newgard, C.B.; et al. Cord Blood Metabolomics: Association With Newborn Anthropometrics and C-Peptide Across Ancestries. *J. Clin. Endocrinol. Metab.* **2019**, *104*, 4459–4472. [[CrossRef](#)]
38. Scholtens, D.M.; Bain, J.R.; Reisetter, A.C.; Muehlbauer, M.J.; Nodzinski, M.; Stevens, R.D.; Ilkayeva, O.; Lowe, L.P.; Metzger, B.E.; Newgard, C.B.; et al. Metabolic Networks and Metabolites Underlie Associations Between Maternal Glucose During Pregnancy and Newborn Size at Birth. *Diabetes* **2016**, *65*, 2039–2050. [[CrossRef](#)]
39. Halket, J.M.; Przyborowska, A.; Stein, S.E.; Mallard, W.G.; Down, S.; Chalmers, R.A. Deconvolution gas chromatography/mass spectrometry of urinary organic acids: Potential for pattern recognition and automated identification of metabolic disorders. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 279–284. [[CrossRef](#)]
40. AMDIS. Available online: <http://www.amdis.net> (accessed on 22 June 2021).
41. Kind, T.; Wohlgemuth, G.; Lee Do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **2009**, *81*, 10038–10048. [[CrossRef](#)]
42. Nodzinski, M.; Muehlbauer, M.J.; Bain, J.R.; Reisetter, A.C.; Lowe, W.L., Jr.; Scholtens, D.M. Metabomxtr: An R package for mixture-model analysis of non-targeted metabolomics data. *Bioinformatics* **2014**, *30*, 3287–3288. [[CrossRef](#)] [[PubMed](#)]
43. Reisetter, A.C.; Muehlbauer, M.J.; Bain, J.R.; Nodzinski, M.; Stevens, R.D.; Ilkayeva, O.; Metzger, B.E.; Newgard, C.B.; Lowe, W.L., Jr.; Scholtens, D.M. Mixture model normalization for non-targeted gas chromatography/mass spectrometry metabolomics data. *BMC Bioinform.* **2017**, *18*, 84. [[CrossRef](#)] [[PubMed](#)]
44. Tyrrell, J.; Richmond, R.C.; Palmer, T.M.; Feenstra, B.; Rangarajan, J.; Metrustry, S.; Cavadino, A.; Paternoster, L.; Armstrong, L.L.; De Silva, N.M.; et al. Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. *JAMA* **2016**, *315*, 1129–1140. [[CrossRef](#)] [[PubMed](#)]
45. Cornelis, M.C.; Agrawal, A.; Cole, J.W.; Hansel, N.N.; Barnes, K.C.; Beaty, T.H.; Bennett, S.N.; Bierut, L.J.; Boerwinkle, E.; Doheny, K.F.; et al. The Gene, Environment Association Studies consortium (GENEVA): Maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet. Epidemiol.* **2010**, *34*, 364–372. [[CrossRef](#)]
46. Das, S.; Forer, L.; Schönherr, S.; Sidore, C.; Locke, A.E.; Kwong, A.; Vrieze, S.I.; Chew, E.Y.; Levy, S.; McGue, M.; et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **2016**, *48*, 1284–1287. [[CrossRef](#)]

47. Taliun, D.; Harris, D.N.; Kessler, M.D.; Carlson, J.; Szpiech, Z.A.; Torres, R.; Taliun, S.A.G.; Corvelo, A.; Gogarten, S.M.; Kang, H.M.; et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **2021**, *590*, 290–299. [[CrossRef](#)]
48. Zheng, X.; Levine, D.; Shen, J.; Gogarten, S.M.; Laurie, C.; Weir, B.S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **2012**, *28*, 3326–3328. [[CrossRef](#)]
49. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **2006**, *1695*, 1–9.
50. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [[CrossRef](#)]
51. Hayes, A.F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*; Guilford: New York, NY, USA, 2013.
52. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]
53. Bojanowski, M. *Intergraph: Coercion Routines for Network Data Objects*; R package version 2.0-2. 2015. Available online: <http://mbojan.github.io/intergraph/> (accessed on 28 April 2022).
54. Butts, C.T.; Leslie-Cook, A.; Krivitsky, P.N.; Bender-deMoll, S. *NetworkDynamic: Dynamic Extensions for Network Objects*; R package version 0.10.0. 2019. Available online: <https://cran.r-project.org/web/packages/networkDynamic/index.html> (accessed on 28 April 2022).
55. Butts, C.T. network: A Package for Managing Relational Data in R. *J. Stat. Softw.* **2008**, *24*, 1–36. [[CrossRef](#)]
56. Butts, C.T. *network: Classes for Relational Data. The Statnet Project*; R package version 1.13.0.1. 2015. Available online: <https://cran.r-project.org/web/packages/network/index.html> (accessed on 28 April 2022).
57. Bender-deMoll, S. *ndtv: Network Dynamic Temporal Visualizations*, R package version 0.13.0. 2019. Available online: <https://cran.r-project.org/web/packages/ndtv/index.html> (accessed on 28 April 2022).