Title

Metabolomics and Genomics in Extremely Premature Infants: Insights into Metabolic Dysregulation and Ancestry-Linked Genetic Variation

ISBN

9798190900217

Author

Malave-Mendez, Thaybeth I.

Publication Date

2026-09-03

Metabolomics and Genomics in Extremely Premature Infants: Insights into Metabolic Dysregulation and Ancestry-Linked Genetic Variation

by
Thaybeth I. Malave-Mendez

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Signed by:
Dara Torgerson
FAE905AA90C3481...

Dara Torgerson
Chair

Signed by:
John A. (Tony) Capra
Date Signed by:

John A. (Tony) Capra

Ryan Hernandez
Date Signed by:

Ryan Hernandez

Akin Oni-Orisan
D2B5F8217A0341B...

Akin Oni-Orisan

Committee Members

Dedication

To my parents, Thallia I. Méndez Figueroa and Iván O. Malavé De Jesús, to whom I owe everything I have accomplished. To all the women in my family who came before me, for whom even dreaming of pursuing higher education or leaving the Island on their own to follow a dream was not a possibility. Finally, to my beautiful island of Puerto Rico, who saw me grow and molded me to become the human I am today.

Dedicatoria

A mis padres, Thallia I. Méndez Figueroa e Iván O. Malavé De Jesús, a quienes les debo todo lo que he logrado. A todas las mujeres de mi familia que vinieron antes que yo, para quienes incluso soñar con cursar estudios superiores o dejar la Isla por su cuenta para perseguir un sueño no era una posibilidad. Por último, a mi hermosa isla de Puerto Rico, que me vio crecer y me moldeó para convertirme en el ser humano que soy hoy.

Acknowledgements

I would not have been able to complete this degree without the support of my mentors, friends, and family.

Thanks to my thesis advisors, Dara Torgerson and Ryan Hernandez, for their passion, patience, dedication, and willingness. Thank you for mentoring with such understanding and compassion, and with the genuine hope that everyone you train grows to be as knowledgeable and successful as you are. You are both the thesis advisors I did not know I needed. I admire you both as a team and as individuals, and I am grateful for the humility with which you approach both science and mentorship.

To Dara, I will never forget the first time I met you in person, during the summer of 2022. You led our conversation by telling me, "I know you might feel lost right now, and that you are improvising your way through this learning process, but do not worry. The truth is that all of us, even at my level, improvise, because that is how science and life work. What happens is that we have more experience, and when we suggest 'solutions,' it may seem like we know everything, but we do not. We are all in the same place, just at different stages." I had never felt so seen and understood during graduate school. Thank you for being vulnerable with me and for always treating me as your equal. To Ryan, beyond thanking you for teaching me to approach science with pure curiosity, I am grateful for all the work you do outside the lab, particularly for communities that have historically been excluded from science and academia. Whenever I dreamed about an ideal environment to pursue my PhD, I imagined being in a space just like this one, where the science is rigorous, students are valued as humans, and where the work being done matters beyond the science itself.

To my current and former lab mates, Steven, Isobel, Sarina, Marisol, Etenesh, Jacqueline Robinson, Melissa, Miguel, and Jackie. You have all taught me so much. Navigating science with all of you has been fascinating. Thanks for inspiring me with your own journey and for always providing feedback and ideas when I needed them most. To my friend Miguel, you are by far one of the most patient humans I know. I felt that patience firsthand as you taught me to code in Python, helped me navigate lab data, and so much more. Being in graduate school alongside you taught me so much about myself, about who I want to become, and, without a doubt, how to be shameless (although this is still a work in progress). To Jackie, although you joined the lab in the latter half of your PhD, it was one of the most remarkable events of my PhD. Having you alongside me to navigate the second half of my own PhD has been so healing. Although we are so different, how alike we are in the way we navigate science (though we are several fields apart) makes my heart so happy. You taught me so much about delegating, working in teams, and pursuing my scientific passion projects.

To my friends whom I met along this journey, thank you for accompanying me and for being my "village" when I needed it most. I want to especially thank my friends Laura, Allison, and Javier. Laura, thank you for being so vulnerable throughout our PhDs. Having someone go through this journey alongside me has been one of the greatest sources of support I have had during this process. Thank you for being my roommate during every QBC retreat; it is literally one of the absolute best parts of my entire PhD. To Allison, thank you for always looking out for me and for being the friend I did not know I needed so much. Thank you for making me part of your life and for making me feel special and loved. It is strange to think that I cannot remember graduate school without you, even though you arrived during the second half of my PhD. You have been a ray of sunshine in my life at moments when I perhaps did not even know I needed a friend. Javier, our friendship was born out of a coincidence, but it became a foundational part of my PhD. From advising me during lab rotations and teaching me how to use the command line,

to reuniting after several years and travelling together, you have been one of the biggest supports during the past two years. Thanks for supporting me inside and outside of science, and listening to me endlessly. To all of you, I am grateful for your friendship and your presence in my life.

To my friends who have been with me since before I started this journey, thanks for staying for the whole ride. To those who were completing their PhDs at the same time as me, Gabriela and Cristian, thanks for commiserating during this marathon with me, even when we were several thousand miles apart. Thanks, Gabriela, for always welcoming me to Michigan and giving me the space to join you as an honorary member of your thesis lab. Your friendship and support were crucial throughout this journey, from being lab partners to being PhD-level scientists.

To my extended family, Fico, what a ride. Even though I am writing all this, I have been tempted more than once to just write a modification of what a great Puerto Rican icon would say. Thanks for not letting me settle when I was deciding on which graduate school to join for my PhD and for supporting me in your own unique way. Thank you for being there and making me part of your life, even from a distance and while navigating your own academic path. Beyond that, I want to express my gratitude to you for welcoming me into your family. Their support was also key during my journey. Particularly, I want to thank your parents and your siblings (including Chelsi) for always checking in on me and making sure I was okay. I do need to give a special shout-out to the youngest of your siblings, Luis, and his best friend, Jaeli (both of whom are also my friends), and thank them for being my virtual desk buddies during a chunk of my PhD thesis writing. Having you both during some of my writing sessions relieved my stress and made me laugh, in ways I would not be able to explain, but beyond that, you both accompanied me through this journey. Thank you both for being there to support me, even if that meant speaking nonsense for hours.

Not even in my wildest dreams could I have imagined I would be in the position I am in today, finishing my PhD at one of the most prestigious biomedical research institutions in the

world. Before I even knew what it meant to dream, my parents learned to dream for me and for everything I was capable of achieving. Throughout my childhood, my dad would tell me over and over, "You will see, when those big-name universities see who you are and talk to you, they are all going to want you to attend their schools, no matter what you decide to study." And he was not wrong, and this was when I was maybe 9 or 10 years old. I have a thousand stories like that one, memories of the countless conversations I had with my parents throughout my childhood and adolescence that always let me keep dreaming about becoming who I am today. Thank you, Papi and Mami, for dreaming and for teaching me to dream, and for always seeing potential in me that, even now, I still cannot quite see in myself.

My parents' motto for my sisters and me has always been, "The most valuable inheritance I can leave you is your education," and there has not been a single second in my life where I have doubted that this is true. So much so that, even after spending their whole lives researching the best schools near our house, the moment my mom realized I had an interest in science, she moved heaven and earth to get me transferred to a specialized science school, even though it was a huge logistical inconvenience for them. That school changed my life and gave me the foundation for the scientist I am today. For that, I will be grateful to them forever, for paying attention to my interests, for sacrificing personally and professionally to support me and my future, and for recognizing my passion for science before I even identified it myself.

I could dedicate an entire chapter to my parents to put in writing all the reasons I owe them my gratitude, because without them, I definitely would not have been able to fulfill this dream. Thank you for being my rocks and my number one fans. Thank you because in the middle of the COVID-19 pandemic, when I found myself with many offers to complete my PhD (which to this day I feel very privileged about), the conversation was never "you should go to this place," but rather "you should go wherever you want" and "if that is where you want to go, how can we help you get there." Thank you for never letting go of me, even while being more than 3,600 miles

away. Thank you for all your support; words will never be enough. Being able to say I am your daughter is a privilege.

To my sisters, the most valuable gift that Mami and Papi have given me, thank you for being my role models and number one cheerleaders. To Thay Liz, the official secretary of the Malavé Méndez family, thank you for reading almost every scientific document and presentation I have ever made, from the personal statements in my graduate school applications to the acknowledgments of this doctoral thesis. Thank you for listening to me give presentations over and over again because "practicing when someone is listening is easier." Thank you because almost daily for the past 5 years, you have listened to me unconditionally every night, especially about the endless complaining that this journey has been. To Thay An, having a sister who has completed a PhD has been one of the most reassuring experiences about this process. I thank God and life for giving me the privilege of having someone so close to me who has completed a parallel degree. Thank you for giving me words of encouragement and peace in many moments when perhaps only you could truly understand what I meant during this academic journey. To Thallia, my "bestie" by choice, thank you for coming to visit me, without expectations, and with a suitcase full of excitement, happiness, love, understanding, and support throughout. Thank you for always paying attention and for being curious about understanding what my scientific projects were about. Last, but definitely not least, thank you for your complicity and for never saying no, even when in our adolescence that meant driving out to pick me up from school, no matter how many hours it took because of traffic, all to support my aspirations. Collectively, I am grateful for the unique and unconditional support each of you gave me throughout this process.

To my family, even though five years later, you still may not be able to explain exactly what I study, you always talk about it with your heads held high and extremely proud of what I do. In particular, I would like to thank my grandparents, Sonia, Octavio, and Ivette. Abuelo Grande and Abuela Sonia, although only one of you will be able to read this, I want to leave in writing the

gratitude I have for you throughout the course of my academic training. Through your respective academic and personal paths, both of you were an example of everything I was capable of achieving, and it was always a recurring conversation in my childhood that, alongside my parents, taught me to dream big. To my Abuela Chínguili, thank you for your curiosity and pride when talking about what I study. You supported Mami as she became the first in our family to attend college and pursue a master's degree, planting the seeds of a legacy that, generations later, led me here. I know that I live and experience a life that you may have never imagined was possible.

Thanks to all the versions of myself that have guided me throughout this journey over the past five years and, in one way or another, have shown up every day along the way. I am definitely not the same person who started this PhD five years ago, but if there is one thing I am sure of, it is that the Thaybeth of 2021 would be extremely proud of the Thaybeth who completed this degree. It has not been an easy journey. There have been many challenges that have required patience, determination, commitment, and discipline, but you were able to overcome every one of them. Thank you for not giving up during the biggest challenge you have faced to date. Read this and remember how capable you are of doing hard things, even when they are filled with uncertainty. Last but not least, I am grateful to God, the universe, and life for the privilege of being able to complete this chapter successfully, and I carry my head high and my heart full of excitement for everything that is yet to come.

Agradecimientos

No hubiese podido completar este doctorado sin el apoyo de mis mentores, amigos y familiares.

Agradezco a mis directores de tesis, Dara Torgerson y Ryan Hernandez, por su pasión, paciencia, dedicación y disposición. Gracias por enseñar desde el entendimiento y la comprensión, sin prejuicios y con ánimos de que todos a los que entrenan lleguen a ser igual o más conocedores y exitosos que ustedes. Ambos son los directores de tesis que no sabía que necesitaba. Los admiro como equipo, pero también en su carácter individual, y agradezco su humildad al navegar la ciencia y cómo ser mentores.

A Dara, nunca olvidaré la primera vez que te conocí en persona en el verano del 2022. Me dijiste “Sé que quizás te sientes perdida ahora mismo y que estás improvisando en este proceso de aprender, pero no te preocupes. La verdad es que todos, incluso a mi nivel, improvisamos, pues así funciona la ciencia y la vida. Lo que pasa es que tenemos más experiencia y cuando les sugerimos “soluciones” parece que lo sabemos todo, pero no. Estamos todos en las mismas, solo en diferentes etapas.”, y nunca me había sentido tan vista y entendida durante este proceso. Gracias por vulnerabilizarte constantemente, y por siempre tratarme como tu igual. A Ryan, aparte de agradecerte por tus enseñanzas desde la curiosidad, agradezco todo el trabajo que haces por las comunidades históricamente excluidas de la ciencia y la academia. Siempre que soñaba sobre un ambiente ideal para hacer mi PhD imaginaba estar en un espacio tal y como este donde la ciencia es rigurosa, el valor por los seres humanos es primordial y donde el trabajo que se hace tiene un impacto que va más allá de la ciencia inmediata.

A mis compañeros de laboratorio, Steven, Isobel, Sarina, Marisol, Etenesh, Jacqueline Robinson, Melissa, Miguel y Jackie. Todos ustedes me han enseñado tanto. Navegar la ciencia

junto a ustedes ha sido muy “top”. Gracias por inspirarme con su propio camino y por siempre ofrecerme “feedback” e ideas cuando más las necesitaba. A mi amigo Miguel, eres, por mucho, uno de los seres humanos más pacientes que conozco. Presencié tu paciencia de primera mano cuando me enseñaste a programar en Python, me ayudaste a navegar la data del laboratorio, y mucho más. Estar en escuela graduada junto a ti me enseñó tanto sobre mí misma, sobre quién quiero llegar a ser, y, sin duda, a tener menos vergüenza (aunque esto sigue siendo un trabajo en progreso). A Jackie, aunque te uniste al laboratorio en la segunda mitad de tu PhD, fue uno de los eventos más memorables de mi PhD. Tenerte a mi lado para navegar la segunda mitad de mi propio PhD ha sido tan sanador. Aunque somos tan diferentes, lo parecidas que somos en la forma en que navegamos la ciencia (aunque estamos en diferentes campos científicos) me hace demasiado feliz. Gracias por enseñarme a delegar, trabajar en equipo, a que pedir ayuda no significa que no puedo sola y a perseguir los proyectos científicos que más me apasionan.

A mis amigos que conocí durante este camino, gracias por acompañarme y ser mi “village” cuando más lo necesité. Quiero agradecer especialmente a mis amigas Laura y Allison, y mi amigo Javier. Laura, gracias por ser tan vulnerable durante el transcurso de nuestros PhD. El tener a una persona atravesando este camino de forma paralela y simultánea a mí ha sido de las fuentes de apoyo más grandes que he tenido durante este proceso. Gracias por ser mi compañera de cuarto durante todos los “QBC retreat”, es literalmente de lo más “top” de todo mi PhD. Tu complicidad y forma de navegar la vida me han enseñado más de lo que mis palabras serán capaces de describir. A Allison, gracias por siempre estar pendiente de mí y ser la amiga que no sabía que tanto necesitaba. Gracias por hacerme parte de tu vida y hacerme sentir especial y querida. Es raro pensar que no recuerdo escuela graduada sin ti, aun cuando llegaste dos años después que yo. Has sido luz en mi vida en momentos que quizás ni sabía que necesitaba una amiga. Javier, nuestra amistad nació de una coincidencia, pero fue una parte fundamental de mi PhD. Desde aconsejarme durante mis rotaciones de laboratorio y enseñarme

a usar el “command line”, hasta reencontrarnos después de varios años y viajar juntos, has sido uno de mis mayores apoyos durante los últimos dos años. Gracias por apoyarme dentro y fuera de la ciencia, y por escuchar mis conversaciones eternas. A cada uno de ustedes, les agradezco su amistad y su presencia en mi vida.

A mis amigos que han estado conmigo desde antes de comenzar este camino, gracias por quedarse durante todo el recorrido. A quienes estaban completando sus PhD al mismo tiempo que yo, Gabriela y Cristian, gracias por sufrir conmigo durante este maratón, aun cuando estaba a miles de millas de distancia. Gracias, Gabriela, por siempre recibirme en Michigan y darme el espacio de unirme a ti como miembro honorario de tu laboratorio de tesis. Tu amistad y apoyo fueron fundamentales a lo largo de este camino.

A mi familia extendida, Fico, wow. Aunque estoy escribiendo todo esto, he sentido la tentación más de una vez de simplemente escribir una modificación de lo que diría un gran prócer boricua. Gracias por no dejarme conformarme cuando estaba decidiendo a qué escuela graduada unirme para mi PhD, y por apoyarme a tu manera única. Gracias por estar ahí y hacerme parte de tu vida, aun en la distancia y mientras navegabas tu propio camino académico. Más allá de eso, quiero expresarte mi gratitud por acogerme dentro de tu familia. Su apoyo también fue clave durante este proceso. En particular, quiero agradecerle a tus papás y a tus hermanos (incluyendo a Chelsi) por siempre estar pendientes de mí y asegurarse de que estuviera bien. Tengo que darle una mención especial y agradecimiento al menor de tus hermanos, Luis, y a su mejor amiga, Jaeli (ambos también amigos míos), por ser mis “desk buddies” virtuales durante gran parte de la redacción de mi tesis doctoral. Tenerlos a ambos durante algunas de mis sesiones de escritura alivió mi estrés y me hizo reír, de formas que no podría explicar, pero, más allá de eso, también me acompañaron a lo largo de este camino. Gracias por apoyarme, aunque eso significara hablar estupideces por horas.

Ni en mis sueños más locos me hubiese imaginado que estaría en la posición en la que estoy hoy, de completar mi PhD en una de las instituciones más prestigiosas en investigación biomédica del mundo. Antes de yo saber lo que era soñar, mis padres aprendieron a soñar por mí y todo lo que yo era capaz de alcanzar. Recurrentemente a través de mi infancia mi papá me decía “ya verás, cuando esas universidades de renombre vean quién tú eres y hablen contigo, todas te van a querer y para lo que sea que quieras estudiar”, y no se equivocó, y todo esto cuando apenas quizás tenía 9 o 10 años. Como esa tengo mil anécdotas que podría escribir sobre el sinnúmero de conversaciones que tuve con mis padres a través de mi infancia y adolescencia, que permitieron que siempre pudiese soñar en convertirme en lo que soy hoy. Gracias papi y mami por soñar y enseñarme a soñar, y por siempre ver el potencial en mí que al sol de hoy no soy capaz de ver.

El lema de vida de mis padres para conmigo y mis hermanas siempre ha sido “la herencia más valiosa que puedo dejarles es su educación”, y no ha habido un segundo en toda mi vida donde haya dudado de que en efecto eso es así. Tanto así, que aun cuando mis padres habían estado toda una vida identificando las mejores escuelas en el perímetro de nuestra casa, tan pronto mi mamá se dio cuenta que yo tenía interés por la ciencia hizo lo imposible para transferirme a una escuela especializada en ciencias, lo que era un total inconveniente para ellos. Esa escuela me cambió la vida y proveyó la base de la científica que soy. Por esto, les daré gracias toda la vida, por estar pendientes de mis intereses, por sacrificarse personal y profesionalmente para cumplir conmigo y mi futuro, e identificar, antes que yo, que me apasionaban las ciencias.

Podría dedicar un capítulo entero a mis padres para dejar plasmado por escrito todas las razones por las que debo agradecerles, porque sin ellos definitivamente no hubiese podido cumplir este sueño. Gracias por ser mi roca y mis fanáticos #1. Gracias porque cuando en el medio de la pandemia del COVID-19 me topé con muchas ofertas para completar mi PhD, la

conversación nunca fue “deberías ir a este sitio”, sino “deberías ir a donde tú quieras” y “si ahí es donde quieres ir, cómo puedo ayudarte a lograrlo”. Gracias por nunca soltarme aun estando a más de 3,600 millas de distancia. Gracias por todo su apoyo, las palabras nunca serán suficientes. Poder decir que soy su hija es un privilegio.

Aparte de un agradecimiento conjunto, quisiera agradecer a mis padres individualmente. A mi papá, gracias por tus llamadas insistentes por “facetime”. Aunque en retrospectiva quizás fueron más las que no contesté que las que sí, más veces de las que piensas fueron un rayito de luz en mis días. Gracias por ser un papá paracaídas y cada vez que pasaba cualquier inconveniente, así fuese lo más mínimo, ver cómo me podías ayudar a resolver. A mi mamá, gracias por ser mi inspiración más grande de las mujeres por las que lucho cada día. Gracias por escuchar cada una de mis anécdotas e inquietudes durante este proceso, y por hacerme sentir que me entendías, aunque al final de la conversación nos riéramos y me dijeras “la verdad es que no entendí el tema como tal, pero mínimo sonaba como que lo comprendes muy bien.”

A mis hermanas, el regalo más valioso que mami y papi me han dado, gracias por ser mi referente y mis “cheerleaders” número uno. A Thay Liz, la secretaria oficial de la familia Malavé Méndez, gracias por leer casi todos los documentos y presentaciones científicas que he hecho, desde los “statements” de mis solicitudes de escuela graduada, hasta la introducción de esta tesis doctoral. Como una oda a nuestra trayectoria de retroalimentación sobre mis escritos y al cierre de esa etapa, te haré llegar los agradecimientos antes de su publicación para que los revises, aunque borre esta oración (o quizás no). Gracias por escucharme dar presentaciones una y otra vez porque “practicar cuando alguien me escucha es más fácil”. Gracias porque casi a diario por los pasados 5 años, me has escuchado todas las noches incondicionalmente, sobre todo cuando hablaba de la letanía que ha sido este trayecto. A Thay An, el tener una hermana que haya completado un PhD ha sido de los aspectos más reconfortantes de este proceso. Agradezco a Dios y a la vida, por darme el privilegio de tener una persona tan cercana que haya

completado un grado paralelo. Gracias por darme palabras de aliento y paz en muchos instantes en los que quizás solo tú podías entender a lo que me refería durante este proceso de formación académica. A Thallia, mi “bestie” por elección, gracias por venir a visitarme, sin expectativas, y con una maleta llena de emoción, felicidad, amor, comprensión, y apoyo durante todo este proceso de completar mi doctorado. Gracias por siempre estar pendiente y tener curiosidad por entender y saber de qué se trataban mis proyectos científicos. Último, pero definitivamente no menos importante, gracias por tu complicidad y por nunca tener un no aun cuando en nuestra adolescencia eso significaba irme a buscar a la escuela y que te tomara horas por el tráfico, todo por apoyar mis aspiraciones. Colectivamente, agradezco el apoyo único e incondicional que cada una me brindó a través de este proceso.

A mi familia, en general, gracias porque aun cuando 5 años después posiblemente todavía no pueden explicar qué estudio, siempre lo llevan con la frente en alto y orgullosos de lo que hago. En particular quisiera agradecerle a mis abuelos, Sonia, Octavio e Ivette. Abuelo Grande y abuela Sonia, aunque solo uno de ustedes podrá leer esto, quiero dejar por escrito el agradecimiento que tengo para con ustedes durante el trayecto de mi formación académica desde mi niñez. Con sus respectivas trayectorias académicas y personales, ambos fueron ejemplo de todo lo que yo era capaz de alcanzar, y siempre fue una conversación recurrente en mi niñez, que, de la mano de mis padres, me enseñó a soñar en grande. A mi abuela Chínguili, gracias por tu curiosidad y orgullo al hablar de lo que hago y a lo que me dedico. Apoyaste a Mami mientras se convertía en la primera de nuestra familia en ir a la universidad y hacer su maestría, sembrando las semillas de un legado que, generaciones después, me trajo hasta aquí. Sé que vivo y experimento una vida que quizás nunca imaginaste como posible.

Agradezco a todas las versiones de mí que me han llevado por este camino y que de una forma u otra han dicho presente cada día de esta trayectoria. Definitivamente no soy la misma que empezó este doctorado hace 5 años, pero si de algo estoy segura es que la Thaybeth del

verano del 2021 estaría muy orgullosa de la Thaybeth que finalizó esta etapa. No ha sido un camino fácil, ha habido muchos retos que han requerido paciencia, determinación, compromiso y disciplina, pero sobrepasaste cada una de esas batallas. Gracias por no darte por vencida en el reto más fuerte que has enfrentado hasta el día de hoy. Lee esto y recuerda que eres capaz de hacer cosas difíciles y llenas de incertidumbre. Por último, pero no menos importante, agradezco a Dios, el universo y la vida por el privilegio de poder finalizar esta etapa con éxito, y voy con la frente en alto y emoción por todo lo que ha de venir.

Para finalizar como diría el gran prócer boricua, Benito Antonio Martínez Ocasio, “un aplauso pa’ mami y papi porque en verdad rompieron.”

Metabolomics and Genomics in Extremely Premature Infants: Insights into Metabolic Dysregulation and Ancestry-Linked Genetic Variation

Thaybeth I. Malavé-Méndez

Abstract

Extremely premature infants exhibit substantial biological heterogeneity that contributes to differences in neonatal outcomes, including bronchopulmonary dysplasia (BPD), one of the most common complications among this population. However, the biological processes underlying this heterogeneity remain incompletely understood. While omics approaches provide powerful tools for investigating these processes, their application in extremely premature infants remains limited. This dissertation used metabolomic and genomic approaches to investigate the urinary metabolome of extremely premature infants at high risk of developing BPD by identifying metabolic pathways associated with BPD and determining how genetic variation captured by genomic ancestry contributes to variation in metabolite levels. Metabolome-wide association analyses during the second and fourth weeks of life identified suggestive metabolite associations with BPD that were enriched for metabolic pathways involved primarily in amino acid and fatty acid metabolism. These patterns of metabolic dysregulation varied across time points and feeding modality (parenteral vs enteral nutrition), with most BPD-differential metabolites being at lower urinary levels during the fourth week of life in infants who develop BPD, particularly among those receiving parenteral nutrition. Beyond BPD-associated metabolic dysregulation, genomic analyses demonstrated that ancestry-linked genetic variation contributes to variation in the urinary metabolome of extremely premature infants. African genomic ancestry proportions were

associated with histidine metabolites and enrichment of the nucleotide metabolism pathway. Several metabolites showed nominal associations in both the BPD and genomic ancestry analyses, highlighting the potential contribution of ancestry-linked genetic variation to variability in clinically relevant metabolites. This dissertation demonstrates that the urinary metabolome of extremely premature infants is shaped by both BPD-associated metabolic dysregulation and ancestry-linked genetic variation, highlighting the value of integrating metabolomics and genomics to understand biological heterogeneity in this vulnerable population.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

^{1}H-NMR	Proton Nuclear Magnetic Resonance
B/AA	Black/African American
Beta	Beta coefficient
BPD	Bronchopulmonary dysplasia
CCC	Lin's Concordance Correlation Coefficient
CEU	Utah residents with Northern and Western European ancestry (1000 Genomes)
CI	Confidence interval
COPD	Chronic Obstructive Pulmonary Disease
DNA	Deoxyribonucleic Acid
Da	Dalton
FDR	False Discovery Rate
GWAS	Genome-wide association studies
HAL	Histidine Ammonia-Lyase
KING	Kinship-based relatedness inference software
LD	Linkage Disequilibrium
LOESS	Locally Estimated Scatterplot Smoothing

MAF	Minor Allele Frequency
MEGA	Multi-Ethnic Genotyping Array
MEX	Mexican residents (Mexican Biobank)
mQTL	Metabolite quantitative trait loci
MS	Mass Spectrometry
MSEA	Metabolite Set Enrichment Analysis
MWAS	Metabolome-wide association study
NHW	Non-Hispanic White
NICU	Neonatal Intensive Care Unit
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
OR	Odds Ratio
p	*p*-value
PC	Principal Component
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
PMA	Post menstrual age
PN	Parenteral Nutrition

PROP	Prematurity and Respiratory Outcomes Program
RSS	Respiratory Severity Score
SE	Standard error
SNP	Single Nucleotide Polymorphism
TOLSURF	Trial of Late Surfactant for Prevention of Bronchopulmonary Dysplasia
UHPLC	Ultra-High Performance Liquid Chromatography
VLBW	Very low birth weight
YRI	Yoruba in Ibadan, Nigeria individuals (1000 Genomes)

Chapter 1: Introduction

Definition of Bronchopulmonary Dysplasia

Bronchopulmonary dysplasia (BPD) is the most common consequence of preterm birth, with its incidence and severity increasing as gestational age and birth weight decrease, particularly among extremely premature infants (<29 weeks gestational age).[1,2] Its definition has been revised over time, reflecting the heterogeneity of clinical manifestations and the limited understanding of its pathogenesis and pathophysiology. Although this complex respiratory disease of prematurity was first described in the late 1960s as a form of chronic lung disease in survivors of severe respiratory distress syndrome, now referred to as the "old BPD," it has since been redefined, with the newer form called the "new BPD".[3,4] While the "old BPD" was described as a ventilator- and oxygen-induced lung injury, the "new BPD" is characterized by disruption to alveolarization, leading to persistent abnormalities in gas exchange, airway function, respiratory mechanics, and lung volumes.[4,5] Although multiple definitions for diagnosis have been proposed for this most recent characterization of the disease, they mostly rely on the physiological need for respiratory support or oxygen supplementation during the first weeks of life, with the assessment of any source of respiratory support or oxygen supplementation beyond 36 weeks post-menstrual age (PMA, typically 8-13 weeks after birth) being the most common approach to define the disease.[6–8] To date, there is no universal consensus on a BPD definition, yet one thing remains consistent across all: BPD is a complex chronic lung disease of prematurity, resulting from lung immaturity, with lasting implications for respiratory health throughout life.

Epidemiology of Bronchopulmonary Dysplasia

BPD is the most common and consequential complication among extremely premature infants. Despite advancements in life-support technologies and neonatal care, the global incidence of BPD has remained largely unchanged over recent decades, with reports that BPD

develops in 10% to 89% of extremely premature infants (10-73% in Europe, 18-89% in North America, 18-82% in Asia, and 30-62% in Oceania) based on studies published through 2017.[9] Differences in disease incidence are likely due to population differences, and/or differences in clinical practice, diagnostic criteria, and/or disease definition. More recent country-specific studies confirm that this trend has persisted over the last decade. In the United States, mortality among infants born at 24 to 28 weeks of gestation decreased from 18.1% to 12.4% between 1997 and 2021, while chronic lung disease increased from 33.4% to 43.3%.[10] In Spain, among infants born before 32 weeks of gestation across three epochs from 2008 to 2022, moderate-to-severe BPD incidence increased from 15.9% to 28.5% while mortality decreased from 13.9% to 8.2%.[11] Similarly, in Australia, mortality among infants born at 24 to 28 weeks of gestation decreased from 17% to 6% between 2010 and 2020, while chronic lung disease incidence increased.[12] In Korea, among infants born at 23 to 27 weeks of gestation between 2014 and 2021, BPD incidence increased from 51.2% to 62.7% as mortality decreased from 28.3% to 19.5%.[13] These trends likely reflect improved survival of extremely premature infants who remain at high risk for developing BPD.

This survival has been largely driven by advances in perinatal and neonatal intensive care, including antenatal corticosteroids, surfactant therapy, improved respiratory support (e.g., use of continuous positive airway pressure), and optimized nutritional management, which have substantially increased the survival of these vulnerable infants.[14–17] Although these advances have improved overall survival in this population, BPD remains a major complication of extreme prematurity, as preventive strategies remain limited to the aforementioned interventions aimed at promoting lung maturation and facilitating discontinuation of respiratory support, and no disease-modifying treatments exist once BPD develops.

Clinical and Biological Risk Factors

The development of BPD is influenced by a continuum of risk factors that begin before conception and continue within the first few weeks of life. Although there is a range of antenatal and postnatal exposures that increase disease risk, prematurity is a prerequisite to BPD development, with extreme prematurity heightening risk, because of the structural and functional immaturity of the developing lung. However, not all premature infants will develop BPD, since factors beyond prematurity contribute to disease susceptibility.

Inherently, conditions that lead to premature birth increase the risk of infants developing BPD, yet specific antenatal factors have been linked to an increased risk of BPD development.[18] Clinical chorioamnionitis, premature rupture of membranes, fetal growth restriction, and hypertensive disorder of pregnancy have all been associated with increased risk of BPD development.[18,19] Meanwhile, intrapartum characteristics such as gestational age, birth weight, and male sex also contribute to disease susceptibility.[18,19] Among these, birth weight and gestational age are the strongest predictors of BPD risk, with decreasing gestational age and birth weight associated with increasing disease incidence. Several postnatal clinical factors have also been identified with increased risk of BPD, such as patent ductus arteriosus, respiratory distress syndrome, sepsis, blood transfusion, and extended duration of mechanical ventilation or oxygen supplementation.[19,20] However, the causal contribution of some of these factors remains debated, as they may also reflect the severity of underlying illness, population differences, and/or differences in clinical care. Hence, BPD risk is shaped, in part, by a combination of antenatal exposures, perinatal characteristics, and postnatal clinical factors, as well as genetic factors, highlighting the multifactorial nature of disease development.

Social and Environmental Exposures

Apart from the clinical and biological risk factors that place extremely premature infants at high risk of developing BPD, social and environmental exposures that may correlate with race/ethnicity have also been associated with increased disease susceptibility and severity. Although Black/African American (B/AA) infants have a significantly lower incidence of BPD compared to non-Hispanic White (NHW) infants, the higher rate of preterm birth among B/AA women results in a disproportionately larger burden of BPD within this population.[21–23] Furthermore, among infants who develop BPD, B/AA infants have worse long-term respiratory outcomes, as evidenced by the higher rates of rehospitalizations, greater use of respiratory medications, and more severe respiratory morbidity later.[21] These differences in disease burden are proposed to be driven, at least in part, by socio-environmental factors, including inequalities in access to and quality of neonatal intensive care units.[24,25] Importantly, the consequences of BPD extend beyond early childhood, with an elevated risk of developing wheezing, decreased lung function, asthma, and other respiratory complications throughout life, besides other extensive comorbidities, which include neurocognitive problems and developmental delays.[5]

BPD Pathology and Pathophysiology

BPD is a complex and multifactorial disease with poorly understood pathology and pathophysiology. The hallmark of BPD is the interruption of lung development caused by premature birth, leading to impaired gas exchange and subsequent neonatal lung injury. Normal lung development *in utero* occurs in five stages: embryonic (gestational weeks 5-7), pseudoglandular (gestational weeks 7-17), canalicular (gestational weeks 17-26), saccular (gestational weeks 26-36), and alveolar (36 gestational weeks-3 years after birth).[26,27] Extremely premature infants are born during the late canalicular or early saccular stages of lung development, prior to the completion of alveolar development and the maturation of pulmonary

vasculature. Consequently, these infants are born with structurally and functionally immature lungs, requiring respiratory support (i.e., positive airway pressure ventilatory support or supplemental oxygen) immediately after birth.[28] Some infants undergo continued lung development after birth, improving pulmonary function and allowing weaning from respiratory support; yet a subset of infants remain dependent on respiratory support beyond the expected period of pulmonary adaptation and thus meet the clinical criteria for a BPD diagnosis.

BPD is highly heterogeneous in clinical presentation and later outcomes, suggesting that it represents a spectrum of related disease processes rather than a single pathological entity.[26,29] In fact, infants with the same disease severity may exhibit different predominant pathological features.[29] Although several attempts have been made to characterize BPD endotypes and multiple classifications have been proposed based on underlying pathophysiology and pathology, the lack of consensus to date highlights the need for further research.[29–33] Nonetheless, several recurring pathological features affecting the alveolar, vascular, airway, and interstitial compartments of the lung have emerged across studies.[26] Among the most frequently reported are decreased alveolarization, reduced vascular growth, abnormal arterial remodeling, small airway bronchoconstriction and hyperreactivity, central airway malacia and stenosis, and interstitial and lymphatic dysfunction.[26,28,29,31] Infants may present a combination of these features, yet no universal features are observed in all infants who develop BPD.

These pathological features are associated with a cascade of pathophysiological processes, primarily characterized by oxidative stress and inflammation.[28,34] These have been identified as key pathophysiological drivers of BPD, triggered by forced gas exchange in underdeveloped lungs during mechanical ventilation, oxygen therapy, and infection, as well as by other prenatal and postnatal exposures. Normal atmospheric oxygen exposure entails a hyperoxic environment for premature infants, as they are used to a hypoxic environment *in utero*.[28,34] The transition to the relatively hyperoxic extrauterine environment, combined with exposure to

respiratory support in extremely preterm infants, which is essential for survival, promotes oxidative stress due to increased oxygen exposure and an immature antioxidant defense system.[28] As a consequence, there is an increased inflammatory response that, jointly with the increased oxidation of a range of cellular components (i.e., lipids, structural proteins, enzymes, and nucleic acids), leads to endothelial cell injury and death, epithelial cell damage, and disruption of the alveolar-capillary membrane, causing alveolar edema and impairment of gas exchange.[28] These interconnected processes alter reparative processes in the developing lung, thereby leading to BPD. Beyond the hyperoxic environment experienced by all premature infants, there are additional common pre- and postnatal exposures that can significantly contribute to these pathophysiological features, particularly inflammation, such as sepsis, patent ductus arteriosus, and chorioamnionitis, all of which have been associated with BPD.[26,28] Despite decades of clinical and basic research, significant gaps remain in our understanding of BPD pathogenesis compared to other major respiratory diseases.

Omics Approaches to Understanding Bronchopulmonary Dysplasia

Omics technologies have emerged as a powerful approach to understand the biological mechanisms underlying complex diseases and to identify potential disease biomarkers. These technologies enable the comprehensive systems-level characterization of biological molecules and processes that cannot be captured through traditional molecular approaches alone, while complementing them. The *omics* era has transformed biomedical research and, in part, reframed the fundamental way we ask and answer scientific questions. By providing a global characterization of distinct layers of the biological system in an unbiased way, scientists have been able to generate hypotheses beyond the constraints of previous scientific discoveries.

Since the early 2000s, with the expansion of genomics and the emergence of other *omics* fields, such as metabolomics, transcriptomics, and proteomics, significant scientific discoveries

have been made in biomedical research. Although most studies have focused on adult populations, *omics* approaches have also been used to study neonatal populations and diseases, including BPD.

Genomics

Genomics is the most established *omics* discipline and focuses on the study of the entire genome, rather than individual genes. This approach has been widely used to identify genetic variants associated with disease susceptibility, drug response, and prognosis.[36] For complex diseases, such as BPD, an important question often addressed prior to conducting genetic association analyses is whether disease risk is influenced, at least in part, by genetic factors.

Heritability of BPD

The genetic contribution to BPD development remains under debate. Multiple twin studies have suggested that the disease is heritable: Lavoie *et al.* and Bhandari *et al.* reported a heritable basis for BPD, with heritability estimates of 53% (16%-89%) and 82% (70%-97%), respectively.[37,38] Meanwhile, Parad *et al*. in 2018 reported no evidence of heritability (0%) for BPD using a similar broad-sense heritability approach.[39] These conflicting findings have opened debate within the field about whether genetics contributes to the development of BPD and, if so, to what extent. Importantly, all heritability estimates reported to date have been derived using broad-sense heritability approaches based on twin data. More recently, Guardado et al. (*unpublished*) applied a narrow-sense heritability approach using genetic data from unrelated extremely premature infants at high risk of developing BPD and identified a non-zero heritable component of 43% (18%-68%) for BPD. Collectively, these findings suggest that although the magnitude of the genetic contribution remains uncertain, genetic factors likely play at least some role in BPD development.

Genetic Association Studies of BPD

Genetic association studies have been conducted to identify variants, genes, and gene pathways involved in BPD development.[40–42] These studies have used a variety of statistical approaches and genetic data types to better understand the genetic architecture of this complex disease, typically evaluating extremely premature infants with and without BPD in case-control comparisons.[41] One of the main caveats of genomic studies, and most omics fields, is the need for large sample sizes to be able to identify statistically significant results. As one can imagine, recruiting extremely premature infants for clinical trials and birth cohorts is extremely challenging due to their vulnerability, heterogeneity, and early mortality. This has limited the power of discovery in BPD genetic studies but has provided valuable insights to inform functional studies and future population-based studies.[43] Overall, the results of these studies have been highly informative, yet replication has been limited.

Candidate gene studies are a common approach used by researchers to validate variants and genes that might be implicated in BPD, given the challenges posed by small sample sizes available in BPD cohorts. The framework for these studies consists of analyzing a set of genes or variants selected based on prior literature and/or functional studies and testing their association with disease in a population. In the context of BPD, these typically involve genes implicated in lung development, inflammation, oxidative stress, and/or tissue injury and repair.[41] Thus, these genomic studies are hypothesis-driven, as they select a set of genes or variants *a priori* and test their association with BPD while sacrificing the broader discovery potential of other statistical genetic approaches (e.g., genome-wide association studies [GWAS]). However, this approach increases statistical power in small samples by reducing the number of statistical tests, thereby reducing the multiple testing burden. Still, for a disease like BPD, challenges remain in identifying candidate genes and variants to evaluate, as much about lung development and the heterogeneous pathology and pathophysiology of BPD remains unknown.[41] Several candidate

gene studies have been conducted for BPD, evaluating surfactant-, toll-like receptor-, antioxidant enzyme-, and interleukin receptor-coding genes and variants.[40,41,44] Although some of these have shown positive associations, they have not been replicated in independent cohorts.[40,44] Table 1.1 shows important genes, genetic variants, and biological pathways that have been implicated in BPD from candidate gene studies or identified as potential candidates through genome-wide association, copy number variation, and whole-exome sequencing studies.

Data-driven, hypothesis-free approaches (i.e., GWAS and exome sequencing studies) have also been used to identify variants and genes that contribute to BPD development. While these provide an unbiased approach to identify novel associations with variants and genes, they come with the high burden of multiple testing correction. GWAS and exome sequencing studies require large sample sizes, yet researchers in neonatology usually work with small cohorts (N<1000). This limitation is particularly relevant given that most variants identified through these approaches exhibit small effect sizes, a pattern commonly observed in complex diseases with modest heritability estimates.[45] As a result, the identification of variants with small effect sizes requires substantially larger sample sizes.[45] Consequently, reaching genome-wide significance in BPD studies has remained a challenge.[40–42] While only a single GWAS has reached statistical significance,[46] these results have not been replicated in subsequent GWAS with larger sample sizes and distinct populations, nor have these studies identified novel significant variants or genes.[43,47] Nonetheless, these studies have identified multiple candidates that have further advanced BPD research by informing future population-based and functional studies (Table 1.1).[40–42] Given the limited power of single-variant tests with the small sample sizes available for BPD studies, researchers commonly aggregate variants and genes into pathways, enabling the identification of biologically relevant processes associated with BPD. This approach has identified multiple pathways implicated in BPD development, including known pathways of lung development, repair, and immune/inflammatory processes, as well as novel pathways further

confirmed in functional studies using *in vitro* and animal models.[41,47] However, the absence of significant GWAS findings led to the hypothesis that the genetic contribution to this disease could be mainly driven by rare variants, which motivated the pursuit of exome sequencing studies.

In contrast to GWAS, which primarily evaluates common non-coding genetic variation across the genome (minor allele frequency (MAF) > 0.05 for typical BPD studies), exome sequencing focuses on protein coding regions. These regions constitute only about 2% of the human genome and are subject to increased purifying selection, which removes most deleterious coding variants from the population. The few that persist usually remain at low frequencies (MAF < 0.01) because purifying selection acts to remove deleterious coding variants from the population. As a result, the rare variants that do exist in exons are disproportionately likely to have functional consequences. This made exome sequencing an attractive approach for studying rare variants with potential functional consequences that could be associated with BPD among a subset of neonatal researchers.[40–42] While highly informative, most of these exome sequencing studies have relied on variant prioritization, biological interpretation of rare variants in candidate genes or pathways, and rare variant quantification in cases versus controls, rather than agnostic exome-wide statistical models. Among the studies, three used statistical models, such as burden tests or sequence kernel association tests, to identify gene associations, yet none identified significant genes after multiple testing correction or replicated any associations.[48–50] Similar to GWAS approaches, the use of pathway enrichment analyses to increase statistical power has also led to the identification of novel pathways implicated in BPD pathophysiology, with genes enriched in pathways involved in lung function and structure, including lung development and repair, immune and inflammatory responses, angiogenesis, cell adhesion, and Wnt signaling (Table 1.1).[40] Results from several of these exome sequencing studies have provided findings that complement those from GWAS, identified novel candidates for further study, and left open the question of the extent to which rare variants contribute to BPD development.

Some factors that might be contributing to inconsistencies in single variant and gene-specific results across population-based studies include differences in the BPD definition, clinical and demographic characteristics, populations sampled, clinical care, and sequencing or genotyping array platforms.[42] The absence of statistically significant findings across genetic studies, and the lack of positive replication, has only fueled the need to further unpack the heterogeneity of BPD, as this is one of the key hypotheses for the lack of significance and replication, alongside small sample sizes. This has motivated the integration of additional omics approaches, such as metabolomics, to evaluate downstream molecular phenotypes that may provide insight into BPD heterogeneity, biological mechanisms, and pathophysiology.

Table 1.1. Summary of genetic studies of bronchopulmonary dysplasia in premature infants

Study	Approach	Sample Size (BPD/Control)	Population Description (BPD/Control)	Main Findings
Kwinta et al., (2008)	Candidate gene	181 (118/63) (preterm infants <32 weeks gestational age)	White (Polish)	Identified an association between *VEGF* −460T>C (rs833061) and BPD, with the −460C/C genotype associated with lower BPD risk. No associations were observed for *TGF-β1*, *IGF-1*, *MTHFR*, or *VEGF* +405G>C polymorphisms.
Hadchouel et al., (2008)	Candidate gene	284 (45/239) (preterm infants <28 weeks gestational age)	Race/ethnicity not reported; Infants recruited from two French NICUs	Identified protective associations for *MMP16* polymorphisms rs2664352 and rs2664349 with BPD, implicating extracellular matrix remodeling and lung development in BPD susceptibility.
Mailaparambil et al., (2010)	Candidate gene	155 (47/108) (preterm infants ≤28 weeks gestational age)	White (German)	Identified nominal associations of *TNFα* (rs1799724), *VEGF* (rs699947), and *TLR10* (rs11096955) with BPD. Findings suggested potential roles for inflammatory and angiogenic pathways in BPD susceptibility.
Prencipe et al., (2011)	Candidate gene	91 (25/66) (preterm infants with RDS, ≤34 weeks gestational age)	22/52 White and 3/14 South American or North African	Identified an association between *MIF* (rs755622; −173G>C) and BPD risk, with the high-producing 173*C allele associated with a lower risk of BPD. Increased *MIF* expression in preterm lung tissue and serum suggested a potential role of *MIF* in lung maturation and repair.
Hadchouel et al., (2011)	GWAS	206 (42/161) discovery; 189 (45/144) replication (preterm infants <28 weeks gestational age)	21/75 European and 21/86 African discovery; 30/112 European and 15/32 African replication	*SPOCK2* identified as associated with BPD in two discovery series. rs1245560 showed a suggestive association and was replicated in an independent population.

Study	Approach	Sample Size (BPD/Control)	Population Description (BPD/Control)	Main Findings
Floros et al. (2012)	Candidate gene (6,324 SNPs across 601 candidate genes)	514 (239/275) discovery; 184 (82/102) replication (preterm infants <35 weeks gestational age)	157/173 Non-Hispanic White and 82/102 African American discovery; 82/102 African American replication	Identified significant associations between *IL18RAP* (rs3771150) and *IL18R1* (rs3771171) and BPD in African American infants. Replication was reported for rs3771150.
Wang et al., (2013)	GWAS	1,726 (899/827) discovery; 795 (371/424) replication (VLBW infants born between 25 to 30 gestational weeks)	174/216 Non-Hispanic White, 117/108 African American, 448/460 Hispanic, 74/93 Asian/Pacific Islander, and 14/22 Other discovery; NA replication	No SNP reached genome-wide significance, and previously reported associations did not pass the corrected significance threshold, with no pathways identified as informative.
Hoffmann et al., (2014)	Genome-wide CNV analysis	1,726 (899/827)	174/216 Non-Hispanic White, 117/108 African American, 448/460 Hispanic, 74/93 Asian/Pacific Islander, and 14/22 Other discovery	Identified several suggestive CNV regions (*ZNF826P*, *ASPSCR1*/*LRRC45*/*STRA13*, *CRELD2*/*ALG12*/*PIM3*/*IL17REL*, *POTEA*–*LINC00293*, and *CARD9*/*GPSM1*/*DNLZ*/*SNAPC4*), but none reached genome-wide significance. No difference in CNV burden was observed between BPD cases and controls.
Carrera et al., (2015)	Exome Sequencing	26 premature infants with BPD	Race/ethnicity not reported; Infants recruited in Italian hospitals	Identified several potential candidate genes implicated in BPD: *NOS2*, *MMP1*, *CRP*, *LBP* and the toll-like receptor (*TLR*) family. These results were exploratory, identified through a prioritization analysis framework.
Li et al., (2015)	Exome Sequencing	50 BPD-affected and unaffected twin pairs (preterm infants <30 week gestational age)	Race/ethnicity not reported; Infants from the California newborn screening program in the United States	Identified 258 genes with rare nonsynonymous mutations in infants with BPD. Enriched pathways included collagen organization, epithelial morphogenesis, and Wnt signaling.

Study	Approach	Sample Size (BPD/Control)	Population Description (BPD/Control)	Main Findings
Ambalavanan et al. (2015)	GWAS	751 (428/323) (extremely low birth weight infants, 23–29 weeks gestational age)	186/153 White, 236/167 Black, 6/2 Other race; 82/74 Hispanic ethnicity.	No SNPs reached genome-wide significance, with some suggestive SNPs identified in adenosine deaminase, and *CD44*, among other genes. Pathway analysis identified lung development and repair pathways, including miR-219 targets, phosphorus oxygen lyase activity (adenylate/guanylate cyclase signaling), *CD44*, and *ADARB2*. Gene expression analyses supported roles for miR-219 and *CD44* in BPD susceptibility.
Ahmad et al., (2016)	Genome-wide CNV analysis	19 preterm infants with BPD, 23 preterm infants with no evidence of lung disease and 41 term infants with no evidence of prematurity or lung disease	Race/ethnicity not reported; Infants recruited in a single center in the United States	Identified recurrent CNVs at 11q13.2, 16p13.3, and 22q11.23–q12.1 containing 21 candidate genes, including *CBL*, *PLAYRR*, *RAB36*, *TSGA10IP*, *SHANK3*, *EP300*, *NFAM1*, *MYH9*, *APOL1*, *APOL2*, *APOL3*, *APOL4*, *APOL5*, PKDREJ, *FOXRED2*, and *HDAC10*. Fifteen genes exhibited developmentally regulated lung expression, supporting these loci as candidates for BPD susceptibility.
Mahlman et al., (2017)	GWAS	174 (60/114) discovery; 555 Finnish and 388 non-Finnish replications (premature infants <31 weeks gestational age)	Discovery cohort was all Finnish infants; Three replication cohorts: (1) Finnish infants 115/221, (2) Finnish infants 31/198, (3) Non-Hispanic White infants from France 48/1163 or Canada 50/51, and French African infants 25/51	Strongest suggestive association was near *CRP*, particularly rs11265269. The association was nominally replicated but did not reach genome-wide significance.
Hamvas et al., (2018)	Exome Sequencing	146 (85/61) (preterm infants <29 weeks gestational age with extreme respiratory phenotypes)	52/34 Non-Hispanic White, race/ethnicity was not reported for the remaining infants	No individual gene reached genome-wide statistical significance, and did not replicate previous results. Identified 345 genes with rare variants unique to infants who developed BPD and 292 genes unique to those who did not. PKA, MAPK, and neuregulin/EGFR pathways were identified to be associated with BPD.

Study	Approach	Sample Size (BPD/Control)	Population Description (BPD/Control)	Main Findings
Torgerson et al., (2018)	GWAS and ancestry GWAS	352 (246/106) (preterm infants <29 weeks gestational age)	136/41 Non-Hispanic White, 82/51 African American, and 28/14 Hispanic White	Identified an association of increased African genomic ancestry proportions in infants of maternal Hispanic white race/ethnicity with increased survival without BPD. No SNP reached genome-wide significance, and previously reported associations did not pass the corrected significance threshold. Top association: intron of *NBL1* (rs372271081). Pathway analysis identified enrichment of immune and inflammatory pathways, including granulocyte/agranulocyte adhesion and diapedesis, Toll receptor signaling, and responses to infection and mechanical ventilation.
Hadchouel et al., (2020)	Exome Sequencing	14 (6/8) discovery; 5 additional severe BPD cases used for replication (preterm infants <31 weeks gestational age)	4/4 Non-Hispanic White and 2/4 African discovery; 3 Non-Hispanic White and 2 African replication	Identified rare variants in nine candidate genes (*BIVM*, *NEDD4*, *CPLX4*, *APOB*, *ANK3*, *ARHGEF10*, *DDAH1*, *CCL18*, and *SREBP1*). No rare variant was shared across multiple infants with severe BPD, and none of the candidate variants replicated in an independent cohort, suggesting a polygenic contribution to severe BPD susceptibility.
Dai et al., (2021)	Exome Sequencing	245 (131/114) (preterm infants <32 weeks gestational age)	Race/ethnicity not reported; Infants recruited in a single center in Shanghai, China	Used exome sequencing, gene-burden analyses, and machine-learning models integrating genetic and clinical data to predict BPD development. Defined a 30-gene BPD risk-gene set and a 21-gene severe-BPD risk-gene set, which improved prediction beyond clinical variables alone.

Study	Approach	Sample Size (BPD/Control)	Population Description (BPD/Control)	Main Findings
Akat et al., (2022)	Candidate gene	192 (96/96) (preterm infants <32 weeks gestational age)	Race/ethnicity not reported; Infants recruited in a single center in Turkey	Identified eight SNPs associated with BPD risk at the allele frequency level, involving *MBL2*, *NFKBIA*, *CEP170*, *MAGI2*, *VEGFA*, *CHST9*, and *KLF12*. The *KLF12* rs4883955 and *CHST9* rs9953270 variants were also associated at the genotype level. Suggests a potential role for Wnt-pathway disruption in BPD pathogenesis.
Blume et al. (2022)	Candidate gene (replication of immunity-related SNPs with fine mapping)	948 (278/670) (preterm infants ≤32 weeks gestational age)	Race/ethnicity distribution not reported; Infants of maternal African and Asian race were excluded, infants were categorized into German, non-German European, and Turkish/Middle Eastern	Identified 5 candidate SNPs nominally associated with BPD: *CRP* (rs11265269), *NUAK1* (rs1427793), *SELL* (rs2229569), *VNN2* (rs1883617), and *CHST3* (rs4148913). *ABCA3* (rs45538638) was the only variant significant after multiple testing correction, supporting roles for cell adhesion, inflammatory signaling, and surfactant production in BPD.
Luo et al. (2023)	Exome Sequencing	66 (34/32) (preterm infants <32 weeks gestational age)	Race/ethnicity not reported; Infants recruited in a single center in Chongqing, China	*BIVM* identified as the strongest candidate gene and a novel *DDAH1* mutation reported. Several candidate genes overlapped with previous exome sequencing studies, while enrichment analyses implicated immune, developmental, and metabolic pathways in BPD susceptibility.

Notes: Only results pertinent to BPD are reported in the main findings. Replication cohorts, when included, are reported in the sample size and population description.

Abbreviations: BPD: bronchopulmonary dysplasia, CNV(s): copy number variant(s), GWAS: genome-wide association study, NICU: neonatal intensive care unit, RDS: respiratory distress syndrome, SNP(s): single nucleotide polymorphism(s), VLBW: very low birth weight.

Metabolomics

Metabolomics involves the quantification of hundreds of low molecular mass biochemicals (metabolites; <2000 Da) present in a biological sample (e.g., urine, plasma, cells). Collectively, these metabolites constitute the metabolome, and their comprehensive characterization has emerged as a powerful approach for studying complex diseases. This more recent omics field can simultaneously capture the combined effects of environmental and genetic factors that may impact human physiology, and thereby increase our overall comprehension of the underlying biological mechanisms and pathogenesis of disease.[51] Metabolomics has been increasingly applied in neonatology for disease diagnosis, prognosis, biomarker discovery, and research on complex neonatal diseases.[52] BPD is no exception.[53]

Metabolic Dysregulation in BPD

Metabolic profiles differ between preterm and term infants.[54–56] More recently, evidence has emerged suggesting that preterm infants can also be stratified into distinct metabolic subtypes.[55] This metabolic heterogeneity may contribute to variability in neonatal outcomes and susceptibility to disease, although its relationship with BPD remains incompletely understood. *In vivo* and *in vitro* models have suggested that metabolic dysregulation plays a key role in BPD pathophysiology.[57] Alterations in major metabolic pathways have been evaluated in these models, and results implicate dysregulation of glucose, lipids, and amino acids.[57] Additionally, infants who develop BPD have been reported to have growth failure and elevated metabolic expenditure, suggesting alterations in metabolic processes.[58] Collectively, these findings suggest that metabolic dysregulation is likely implicated in BPD development and progression.

Metabolomic Studies of BPD

Metabolomic studies of BPD have been conducted to characterize metabolic differences between infants with and without BPD using population-based data.[53] This approach enables the identification of metabolites and metabolic pathways associated with disease development, providing insight into the biological mechanisms underlying BPD while also supporting biomarker discovery. Similar to genetic association studies, metabolomic studies within the neonatal population also struggle with statistical power to detect associations due to the small sample sizes of cohorts with biofluids available for metabolomics profiling. However, approaches analogous to those used in genomics have been implemented to maximize statistical power and identification of biological mechanisms that play a key role in BPD development.

BPD metabolomic studies can be broadly categorized into targeted and untargeted approaches, each providing distinct biological insights (Table 1.2 and Table 1.3). Targeted metabolomics is hypothesis-driven and focuses on the quantification of predefined metabolites selected based on prior evidence of biological role, a framework comparable to candidate gene analyses. Building upon evidence from experimental models, BPD targeted metabolomic approaches have primarily focused on the evaluation of a range of amino acids, carnitines, and lipids. These metabolite classes have been prioritized due to their biological relevance to lung development and injury. Acylcarnitines play a central role in mitochondrial fatty acid β-oxidation by transporting long-chain fatty acids into mitochondria for energy production, whereas lipids are key constituents of pulmonary surfactant and contribute to the maintenance of normal respiratory function.[59] Consistent with findings from *in vivo* and *in vitro* models, most of these studies report that certain amino acids (e.g., alanine), carnitines (e.g., medium-chain acylcarnitines), and lipids differ between infants who develop BPD and those who do not.[59–63] Two of these were longitudinal studies that suggest that amino acid and carnitine metabolomic profiles differ by postnatal age[59] and vary with nutritional stages[60] (e.g., maximum amino acid intake during parenteral nutrition

and full enteral nutrition), highlighting the temporal dependence of metabolomics sampling (Table 1.2). However, specific metabolites reported to differ between infants with and without BPD have differed across studies, and replication of individual findings has been limited.[59–64]

Untargeted metabolomics is data-driven and aims to comprehensively characterize the metabolome, enabling the unbiased evaluation of metabolites associated with BPD. Although this approach provides broad coverage of the metabolome, the metabolites identified and quantified may differ depending on the metabolomic platform and processing methods. Untargeted metabolomic studies of BPD have used multiple statistical approaches with the aim to maximize discovery. The majority of these studies have used supervised multivariate classification methods (e.g., PLS-DA and OPLS-DA) to evaluate differences in metabolic profiles between infants with and without BPD, followed by secondary analyses to identify metabolites contributing to differences between groups.[65–70] Despite the exploratory nature of this approach, the number of metabolites characterized across studies that reported this metric ranged from approximately 15 to 300, highlighting substantial heterogeneity in analytical depth and limiting overlap in the metabolites identified across studies. Thus, there has been limited replication of individual results. However, consistent with findings from targeted metabolomic studies of BPD, amino acids and carnitines were among the metabolite classes most frequently reported to differ between infants with and without BPD (Table 1.3).[63,65–67,69–72]

This heterogeneity likely reflects, in part, methodological differences in biofluids analyzed (e.g., dried blood spots versus tracheal aspirates), metabolite panels, and platforms (e.g., ^{1}H-NMR and various MS methods), which limit direct comparisons and replication of findings across studies. Because metabolite profiles are highly dependent on the biological specimen analyzed, biofluid selection represents an important consideration in metabolomic studies. Different biofluids reflect distinct biological processes and therefore contain unique metabolite compositions. Additional factors, such as the timing of sample collection, gestational age, birth weight, nutrition

source, clinical care, and BPD definition, may further contribute to variability in metabolomic profiles and study findings. These factors, along with the small sample sizes of the available cohorts, are among the primary limitations of current BPD metabolomic studies, reducing the generalizability of the findings to the broader neonatal population. Nevertheless, despite substantial methodological and sampling heterogeneity, the overall pattern of findings supports the role of metabolic dysregulation in BPD and demonstrates the value of metabolomics to elucidate disease mechanisms, identify biologically relevant pathways, and inform future biomarker discovery efforts in this population.[53,57,73]

Table 1.2. Summary of targeted metabolomics studies of bronchopulmonary dysplasia in premature infants

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
Piersigilli et al. (2019)	NA; included acylcarnitines, amino acids, biogenic amines, cholines and sphingomelins	68 ventilated preterm infants (gestational age < 32 weeks; 44/24)	Tracheal aspirates	First week of life	UHPLC–QTRAP–MS/MS	PLS-DA and non-parametric test	53 metabolites were identified as characteristic of BPD, with 18 metabolites being highly significant in the separation by BPD status. Among them were both amino acids and acylcarnitines, most of which were higher in infants with BPD.
Wang et al. (2021)	18 amino acids and 40 carnitines	85 preterm infants (gestational age < 32 weeks; 45/40)	Dried blood spots	Four time points depending on feeding transitions*	MS/MS	PLS-DA, univariate analysis and multivariate logistic regression	Amino acids and carnitines metabolic profiles changed between infants with and without BPD, after birth and during feeding transitions. Metabolic profiles varied by time points, and four amino acids and four carnitines differed between groups.
Frazer et al. (2022)	8 short chain fatty acids	72 preterm infants (gestational age < 28 weeks; 45/27)	Stool	Day 14 and 28 after birth	LC-MS/MS equipped with a C18 reversed-phase UHPLC	Non-parametric test and logistic regression	Identified that lower levels of acetic acid in the stool are associated with increased odds of BPD at both time points evaluated.
Ye et al. (2022)	72 amino acids and acylcarnitines	205 preterm infants (27-28 weeks gestational age; 97/108)**	Dried blood spots	Four time pointst	MS/MS	Logistic regression, clustering methods and random forest	27 metabolites were longitudinally associated with BPD, including both metabolite classes evaluated. The direction of association varied over time, and not all metabolites demonstrated significant associations at each time point.
Guo et al. (2024)	86 amino acids and carnitines	136 preterm infants (gestational age < 32 weeks; 38/98)	Dried blood spots	Within the first 24 hours of life	LC-MS/MS	PLS-DA, univariate parametric and non-parametric tests	Observed a reduction of histidine levels, and ornithine/citrulline ratio, and ratios of acylcarnitines C3/C0 and C5/C0 in infants who develop BPD.

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
La Frano et al. (2018)	NA: included biogenic amines and oxylipins	Discovery sample (n=42): preterm infant < 30 weeks gestational age (34/8) Validation sample (n=20): preterm infants < 28 weeks gestational age (20/0)	Umbilical cord blood plasma	Collected at delivery	UHPLC-QTOF-MS; GC–TOF-MS	non-parametric Kendall rank correlations and welch t-test	Secondary analyses identified 47 metabolites associated with BPD and severity at $p<0.05$, including oxylipins, complex lipids, and primary metabolites, which included several amino acids. Only one oxylipin remained significant after multiple testing correction.

Notes: Only results pertinent to BPD were reported in the main findings. None of the studies included a replication cohort for the BPD metabolomic analyses, and most of these studies do not account for multiple-testing correction. *Time points included: <24hrs of life, maximum levels of amino acids in parenteral nutrition before enteral nutrition, ratio of enteral nutrition/(enteral nutrition + parenteral nutrition) energy intake reached about 50%, and full enteral nutrition. †Time points included: within 24 hours of delivery, days 7–8, 27–29, and 42–43. NA indicates that the corresponding information was not reported.

Abbreviations: GC-TOF-MS: gas chromatography–time-of-flight mass spectrometry; LC-MS/MS: liquid chromatography–tandem mass spectrometry; MS/MS: tandem mass spectrometry; PLS-DA: partial least squares discriminant analysis; UHPLC-QTOF-MS: ultra-high-performance liquid chromatography–quadrupole time-of-flight mass spectrometry; UHPLC-QTRAP-MS/MS: ultra-high-performance liquid chromatography–quadrupole linear ion trap tandem mass spectrometry.

Table 1.3. Summary of untargeted metabolomics studies of bronchopulmonary dysplasia in premature infants

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
Fanos et al. (2014)	NA	36 preterm infants (gestational age < 29 weeks; 18/18)	Urine	Within 24-36 hours after birth	^{1}H-NMR	OSC-PLS-DA	Discriminant analysis separated the two groups (BPD vs. No BPD). Lactate, taurine, trimethylamine-N-oxide, myoinositol were increased in infants who developed BPD compared to those who did not, while levels of gluconate were lower.

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
Baraldi et al. (2016)	NA	21 preterm infants (mothers had undergone amniocentesis between 21-32 weeks of gestation; 10/11)	Amniotic fluid	21-28 weeks gestational age during preterm labor	UPLC-MS	oCPLS2-DA, t-test	Suggest prenatal amniotic fluid metabolomic profiling could identify preterm fetuses at risk for developing BPD. Altered fatty acid metabolism, and reduced methylation capacity and antioxidant precursors observed.
Pintus et al. (2018)	15	18 preterm infants (gestational age < 28 weeks; 7/12)	Urine	Day 7 of life	^{1}H-NMR	OPLS-DA	Discriminant analysis separated infants who developed and those who did not develop BPD. Alanine and betanine were increased in infants who developed BPD compared to those without BPD, and trimethylamine-N-oxide, lactate and glycine were lower.
Lal et al. (2018)	NA	30 preterm infants (gestational age <28 weeks; 15/15)	Tracheal aspirates	At birth or within 6 hours of birth before surfactant administration	LC-MS/MS	PCA, PLS-DA and Mummichog for pathway enrichment analysis	Metabolites within the fatty acid oxidation and androgen and estrogen pathways were decreased in infants who develop BPD, compared to those who do not develop the disease. Suggest that the airway microbiome may be altering the metabolome.

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
Course et al. (2023)	235	214 children (born ≤ 34 or ≥ 37 weeks' gestation; 144 were born preterm; 34/110)	Exhaled breath condensates	Children were 7 to 12 years old	GCTOF-MS	Anova and MSEA	10 metabolites were identified to be different between preterm infants with and without BPD and 14 between term and preterm infants with BPD. There was overlap in differential metabolites across comparisons, and amino acids accounted for most of them. Enrichment analyses revealed that urea and glutathione metabolism were altered.
You et al. (2024)	>300	30 preterm infants (gestational age and birth weight matched; 15/15)	Plasma	36 weeks PMA	LC-MS/MS	OPLS-DA, t-test	Identified 10 metabolites that differentiated infants with BPD from those without the disease. Most differential metabolites were amino acids, finding an enrichment with three metabolic pathways.
Chiu et al. (2024)	44	89 infants (48 preterm of 28-32 weeks of gestation and 50 term; 48/50)	Urine	6 months corrected age	1H-NMR	Univariate parametric and non-parametric tests	21 metabolites were differential when evaluating gestational age and 24 with BPD severity. 19 metabolites overlapped across both analyses, and a substantial proportion were amino acids.
Bonadies et al. (2025)	740	50 preterm infants (gestational age and birth weight matched; 25/25)	Urine	Within the first 24 hours of life	UPLC-QTOF-MS	Univariate parametric and non-parametric tests, PLS and logistic regression	Identified 17 metabolites to be implicated in BPD development across analyses, including 4 amino acids and 1 peptide.

Study	Number of metabolites evaluated	Sample Size (BPD/Control)	Sample	Time of Sampling	Platform	Statistical Model	Main Findings
La Frano et al. (2018)	1,656	Discovery sample (n=42): preterm infant < 30 weeks gestational age (34/8) Validation sample (n=20): preterm infants < 28 weeks gestational age (20/0)	Umbilical cord blood plasma	Collected at delivery	UHPLC-QTOF-MS; GC–TOF-MS	non-parametric Kendall rank correlations and welch t-test	Secondary analyses identified 47 metabolites associated with BPD and severity at $p<0.05$, including oxylipins, complex lipids, and primary metabolites, which included several amino acids. Only one oxylipin remained significant after multiple testing correction.

Notes: Only results pertinent to BPD were reported in the main findings. None of the studies included a replication cohort for the BPD metabolomic analyses, and most of these studies do not account for multiple-testing correction. NA indicates that the corresponding information was not reported.

Abbreviations: ANOVA: analysis of variance; GC-TOF-MS: gas chromatography–time-of-flight mass spectrometry; LC-MS/MS: liquid chromatography–tandem mass spectrometry; MSEA: metabolite set enrichment analysis; oCPLS2-DA: orthogonal consensus partial least squares discriminant analysis; OPLS-DA: orthogonal partial least squares discriminant analysis; OSC-PLS-DA: orthogonal signal correction–partial least squares discriminant analysis; PCA: principal component analysis; PLS-DA: partial least squares discriminant analysis; PMA: postmenstrual age; ^{1}H-NMR: hydrogen nuclear magnetic resonance; UHPLC-QTOF-MS: ultra-high-performance liquid chromatography–quadrupole time-of-flight mass spectrometry; UPLC-MS: ultra-performance liquid chromatography–mass spectrometry; UPLC-QTOF-MS: ultra-performance liquid chromatography–quadrupole time-of-flight mass spectrometry.

Metabolomics and Genomics Integration

Metabolomics represents a downstream molecular layer that integrates information from other *omics* technologies, particularly the genome, as well as environmental exposures, providing a direct readout of ongoing physiological processes. Several studies have demonstrated that genetic variation contributes to inter-individual differences in metabolite levels, supporting a heritable component to human metabolic individuality.[74,75] In fact, heritability has been estimated for several metabolite classes, demonstrating that the genetic contribution to metabolite variation differs across metabolite classes.[76] As a result, considerable efforts have been made to identify genetically regulated metabolites that may serve as intermediate phenotypes linking genetic variation to disease.[74]

In respiratory diseases, such as asthma, studies integrating metabolomics and genomics have been performed to identify metabolite quantitative trait loci (mQTL). These have successfully identified several genetically regulated metabolites that also associate with disease status.[77,78] However, to date, integrative metabolomic and genomic studies have not been performed in BPD. From a statistical standpoint, these studies are essentially GWAS where the phenotypes are the metabolite levels. Additional studies using other statistical approaches (e.g., mendelian randomization) have been performed to determine whether these genetically regulated metabolites linked to disease are a cause or a consequence of disease, providing unique insights into disease pathology.[79,80] The integration of these complementary omics provides an unprecedented opportunity to dissect the complex pathophysiological mechanisms underlying disease, as each *omics* layer contributes distinct yet interconnected information that cannot be captured by a single-omics approach alone. These integrative approaches have expanded beyond metabolomics and genomics to incorporate multiple molecular layers (e.g., transcriptomics and proteomics), giving rise to what is now referred to as multi-omics.[36]

Population Diversity in Omics Research

Increasing evidence demonstrates the importance of diverse sampling in omics research to increase the generalizability of findings and clinical utility across populations.[81–83] However, to date, most omics studies, particularly genomics, have been conducted in individuals of European descent.[84] This imbalance has limited our understanding of biological variation across populations and contributed to disparities in the clinical implementation of precision medicine. This disparity has been most evident in genomics, largely because it was the first omics field to emerge and continues to be the most established. Around 96% of the GWAS Catalog participants are of European descent, and this imbalance extends beyond GWAS studies into other omics: 93.15% of participants in published proteome-wide association studies are of European descent.[85] This has also directly impacted BPD research, where the majority of the individuals included in genomics studies of this disease have been of European descent (Table 1.1).

Overall, this imbalance has directly affected precision medicine tools such as polygenic risk scores, which, although not yet developed for BPD, are typically developed in predominantly European populations and often exhibit reduced accuracy and transferability when applied to individuals from other ancestral backgrounds.[83] The gap has been further amplified by the emergence of large biobanks (e.g., UK Biobank), in which the vast majority of participants (94%) are of European ancestry.[85] As these resources are increasingly used for genetic and multi-omic analyses, they continue to reinforce the underrepresentation of diverse, historically excluded populations in biomedical research. However, several large-scale initiatives (e.g., *All of Us* and the *Trans-Omics for Precision Medicine* program) have been established in an effort to improve representation in biomedical research and increase opportunities for discovery across diverse populations.[86,87] A crucial concept underlying these initiatives is genetic ancestry, which has important implications in the architecture of complex traits and the transferability of omics-based findings across populations.

Population Descriptors

The use of population descriptors, including genetic ancestry and race/ethnicity, has received increasing attention in human genetics and genomics research, given that these terms represent distinct concepts yet have sometimes been used inconsistently or interchangeably.[88–92] In response, the National Academies of Sciences, Engineering, and Medicine published a consensus report providing a framework and recommendations for the appropriate and transparent use of population descriptors in genetics and genomics research.[91]

Genetic ancestry may correlate with race/ethnicity, yet these are distinct, non-interchangeable concepts. While genetic ancestry reflects inherited genetic relationships shaped by shared population history, race/ethnicity are sociopolitical constructs commonly based on self-identification and influenced by social, cultural, environmental, and historical factors, as well as perceived physical characteristics.[91] Genetic ancestry, however, can be inferred from patterns of genome-wide genetic similarity,[91] including variation in allele frequencies and linkage disequilibrium. These genome-wide patterns form the basis of several methods developed to estimate an individual's genomic ancestry proportions and locus specific genomic ancestry.[93]

Although human genetic variation is complex and continuous, patterns of genetic variation and inferred genetic ancestry have commonly been summarized using broad continental population descriptors, such as African, European, and Amerindigenous.[91,92] However, the use of broad continental categories has been increasingly questioned because they may inadequately represent the continuous and multidimensional nature of human genetic variation and are sensitive to the reference populations and sampling strategies used.[92] More recent statistical approaches, born of evolutionary theory, are working to move away from these modalities and reconceptualize the patterns of genetic similarity used to infer genetic ancestry as what they are: a continuum that reflects human evolutionary history.[94,95] Nonetheless, genetic ancestry inferred

from patterns of genetic similarity has been well established as a tool for capturing population-level genetic structure and is used both as a methodological tool to account for population structure[96,97] and as a variable for understanding differences in traits.[98,99] However, it is important to note that patterns of human genetic variation evolve with ongoing admixture, migration, and demographic changes, which continue to shape the composition of human populations over time; populations are not static.

Ancestry-linked Traits

These patterns of genetic variation have been linked to differences in a broad range of clinical and molecular phenotypes across populations.[100,101] Several studies have reported associations between ancestry-linked genetic variation and disease susceptibility,[102] and in some contexts, recent genetic shifts have directly impacted the association of these patterns of variation and biomedical traits.[98] Within the context of BPD, there is a GWAS that identified an association of increased African genomic ancestry proportions in infants of maternal Hispanic white race/ethnicity with increased survival without BPD.[103] However, this protective effect was not significant among infants of maternal B/AA race, and these findings have not been replicated in an independent study.

Metabolites can be considered molecular phenotypes. Increasing evidence has demonstrated the contribution of genetic variation to the metabolome,[104] yet the contribution of ancestry-linked genetic variation to metabolite levels remains poorly understood. Nevertheless, a recent study reported that ancestry-linked genetic variation contributes to differences in metabolite levels in admixed individuals, particularly for lipids and amino acids, both of which are frequently implicated in disease.[99] These findings suggest that ancestry-linked genetic variation may contribute to population-level differences in the metabolome and potentially influence molecular pathways relevant to disease. However, the extent to which ancestry-linked genetic

variation contributes to metabolite variation, and whether ancestry-associated metabolite variation contributes to disease susceptibility and progression, remains largely unknown, particularly within the context of BPD.

Justification for the Use of Population Descriptors

Throughout this dissertation, both race/ethnicity and genetic ancestry are used as population descriptors, recognizing that while they may often correlate, they represent distinct concepts.[91] Their use allows for the evaluation of population-level differences relevant to both the epidemiology of BPD and the metabolomic and genomic analyses performed. Race/ethnicity is used to capture socially defined population differences that may be associated with environmental, social, clinical, and structural factors. For the purposes of this dissertation, genetic ancestry refers to ancestry inferred from patterns of genetic similarity between an individual and contemporary reference populations, reflecting aspects of shared population history shaped by processes such as migration and admixture. Genetic ancestry is used to characterize ancestry-linked patterns of genetic variation relevant to the genomic and metabolomic analyses performed in this dissertation. Although broad continental descriptors simplify the continuous and multidimensional nature of human genetic variation, this approach provides an interpretable framework for characterizing broad patterns of genetic variation and evaluating ancestry-associated variation.[91,92] It also allows population diversity to be explicitly evaluated and reported, increasing the visibility of populations that have been historically excluded from biomedical and genomic research. More continuous and multidimensional approaches may better represent human genetic variation.[94,95] As these methods continue to develop, future studies should consider approaches that more fully capture continuous patterns of genetic similarity while maintaining interpretability and analytical feasibility. Race/ethnicity was not used as a proxy for genetic ancestry, nor was genetic ancestry considered a substitute for race/ethnicity for any of the analyses in this dissertation.

Dissertation Overview and Objectives

BPD remains one of the primary complications of extreme prematurity and continues to represent a substantial burden on affected infants and their families despite advances in neonatal care. Although progress has been made in understanding the pathology and pathophysiology of BPD, important gaps remain in our understanding of the biological mechanisms underlying its heterogeneous pathology and long-term outcomes. Part of this limited understanding comes from the challenges of studying extremely premature infants, including small sample sizes, limited sample availability, and the historical lack of genetic diversity in population-based studies.

Evidence from experimental models and population-based metabolomic studies suggests that metabolic dysregulation contributes to BPD development and progression. However, comprehensive studies in larger cohorts with repeated measurements are still needed to further characterize the metabolic pathways associated with disease susceptibility and progression, thereby providing information on biological mechanisms implicated in this disease. Furthermore, while genetic variation has been shown to contribute to metabolite variation, the extent to which ancestry-linked genetic variation contributes to metabolite differences remains poorly understood. This knowledge gap is particularly relevant for metabolites and metabolic pathways associated with disease processes, including BPD.

Thus, by leveraging metabolomics and genomics data from two cohorts of extremely premature infants, this dissertation aims:

1. To identify metabolites and metabolic pathways associated with BPD within the first month of life, prior to disease diagnosis.
2. To determine the contribution of ancestry-linked genetic variation to the differences in metabolite levels in extremely premature infants.

3. To evaluate if ancestry-linked genetic variation contributes to differences in the levels of BPD-differential metabolites.

References

1. Taylor, G. L. & O'Shea, T. M. Extreme Prematurity: Risk and Resiliency. *Curr Probl Pediatr Adolesc Health Care* **52**, 101132 (2022).

2. Jensen, E. A. & Schmidt, B. Epidemiology of Bronchopulmonary Dysplasia. *Birth Defects Res A Clin Mol Teratol* **100**, 145–157 (2014).

3. Northway, W. H., Rosan, R. C. & Porter, D. Y. Pulmonary Disease Following Respirator Therapy of Hyaline-Membrane Disease. *New England Journal of Medicine* **276**, 357–368 (1967).

4. Perveen, S., Chen, C.-M., Sobajima, H., Zhou, X. & Chen, J.-Y. Editorial: Bronchopulmonary dysplasia: latest advances. *Front. Pediatr.* **11**, (2023).

5. Epidemiology of bronchopulmonary dysplasia - Jensen - 2014 - Birth Defects Research Part A: Clinical and Molecular Teratology - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/bdra.23235.

6. Shennan, A. T., Dunn, M. S., Ohlsson, A., Lennox, K. & Hoskins, E. M. Abnormal pulmonary outcomes in premature infants: prediction from oxygen requirement in the neonatal period. *Pediatrics* **82**, 527–532 (1988).

7. Jensen, E. A. *et al.* The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. *Am J Respir Crit Care Med* **200**, 751–759 (2019).

8. Walsh, M. C. *et al.* Impact of a physiologic definition on bronchopulmonary dysplasia rates. *Pediatrics* **114**, 1305–1311 (2004).

9. Siffel, C., Kistler, K. D., Lewis, J. F. M. & Sarda, S. P. Global incidence of bronchopulmonary dysplasia among extremely preterm infants: a systematic literature review. *The Journal of Maternal-Fetal & Neonatal Medicine* **34**, 1721–1731 (2021).

10. Horbar, J. D. *et al.* Trends in Mortality and Morbidities for Infants Born 24 to 28 Weeks in the US: 1997–2021. *Pediatrics* **153**, e2023064153 (2023).

11. Orvay, J. A. C., Baucells, B. J., Aloy, J. F., Echegoyen, X. M. & Sebastiani, G. Trends in bronchopulmonary dysplasia and 15-year evolution in respiratory and hemodynamic management of very premature infants. *Sci Rep* **15**, 24987 (2025).

12. Legge, N., Popat, H. & Fitzgerald, D. Improved survival at the cost of more chronic lung disease? Current management and outcomes in extremely preterm infants born in New South Wales and the Australian Capital Territory: 2010–2020. *World J Pediatr* **20**, 230–238 (2024).

13. Jeon, G. W., Oh, M. & Chang, Y. S. Increased bronchopulmonary dysplasia along with decreased mortality in extremely preterm infants. *Sci Rep* **15**, 8720 (2025).

14. Morgan, A. S., Mendonça, M., Thiele, N. & David, A. L. Management and outcomes of extreme preterm birth. *BMJ* **376**, e055924 (2022).

15. Härtel, C. *et al.* Association of Administration of Surfactant Using Less Invasive Methods With Outcomes in Extremely Preterm Infants Less Than 27 Weeks of Gestation. *JAMA Netw Open* **5**, e2225810 (2022).

16. Sindelar, R., Nakanishi, H., Stanford, A. H., Colaizy, T. T. & Klein, J. M. Respiratory management for extremely premature infants born at 22 to 23 weeks of gestation in

proactive centers in Sweden, Japan, and USA. *Seminars in Perinatology* **46**, 151540 (2022).

17. Koletzko, B., Poindexter, B. B., Domellöf, M. & Cheah, F.-C. Nutritional Care of Preterm Infants : Scientific Basis and Practical Guidelines. 1–460 (2021).

18. Trembath, A. & Laughon, M. Predictors of Bronchopulmonary Dysplasia. *Clin Perinatol* **39**, 585–601 (2012).

19. Xiong, P., Li, L., Yu, Z., Pu, Y. & Tang, H. Risk factors for bronchopulmonary dysplasia in preterm infants: a systematic review and meta-analysis. *PeerJ* **13**, e20202 (2025).

20. Dankhara, N., Holla, I., Ramarao, S. & Kalikkot Thekkeveedu, R. Bronchopulmonary Dysplasia: Pathogenesis and Pathophysiology. *J Clin Med* **12**, 4207 (2023).

21. Ryan, R. M. *et al.* Black Race is Associated with a Lower Risk of Bronchopulmonary Dysplasia. *J Pediatr* **207**, 130-135.e2 (2019).

22. Patel, P., Ellefson, A. & Paul, D. A. Racial Disparities Among Predicted Bronchopulmonary Dysplasia Risk Outcomes in Premature Infants Born <30 Weeks Gestation. *Health Equity* **7**, 825–830 (2023).

23. Janevic, T. *et al.* Association of Race/Ethnicity With Very Preterm Neonatal Morbidities. *JAMA Pediatr* **172**, 1061–1069 (2018).

24. Howell, E. A. *et al.* Differences in Morbidity and Mortality Rates in Black, White, and Hispanic Very Preterm Infants Among New York City Hospitals. *JAMA Pediatr* **172**, 269–277 (2018).

25. Alur, P., Holla, I. & Hussain, N. Impact of sex, race, and social determinants of health on neonatal outcomes. *Front. Pediatr.* **12**, (2024).

26. Thébaud, B. *et al.* Bronchopulmonary dysplasia. *Nat Rev Dis Primers* **5**, 78 (2019).

27. Tagg, A. Bronchopulmonary dysplasia (BPD). *Don't Forget The Bubbles* https://doi.org/10.31440/DFTB.49274 (2022) doi:10.31440/DFTB.49274.

28. Kalikkot Thekkeveedu, R., Guaman, M. C. & Shivanna, B. Bronchopulmonary dysplasia: A review of pathogenesis and pathophysiology. *Respiratory Medicine* **132**, 170–177 (2017).

29. Wu, K. Y. *et al.* Characterization of Disease Phenotype in Very Preterm Infants with Severe Bronchopulmonary Dysplasia. *Am J Respir Crit Care Med* **201**, 1398–1406 (2020).

30. Day, C. L. & Ryan, R. M. Bronchopulmonary dysplasia: new becomes old again! *Pediatr Res* **81**, 210–213 (2017).

31. Pierro, M., Van Mechelen, K., van Westering-Kroon, E., Villamor-Martínez, E. & Villamor, E. Endotypes of Prematurity and Phenotypes of Bronchopulmonary Dysplasia: Toward Personalized Neonatology. *Journal of Personalized Medicine* **12**, 687 (2022).

32. Collaco, J. M. & McGrath-Morrow, S. A. Respiratory Phenotypes for Preterm Infants, Children, and Adults: Bronchopulmonary Dysplasia and More. *Ann Am Thorac Soc* **15**, 530–538 (2018).

33. Moreira, A. G. *et al.* Leveraging transcriptomics to develop bronchopulmonary dysplasia endotypes: a concept paper. *Respir Res* **24**, 284 (2023).

34. Ambalavanan, N., Deutsch, G., Pryhuber, G., Travers, C. P. & Willis, K. A. The evolving pathophysiology of bronchopulmonary dysplasia. *Physiological Reviews* **106**, 197–237 (2026).

35. Shahzad, T., Radajewski, S., Chao, C.-M., Bellusci, S. & Ehrhardt, H. Pathogenesis of bronchopulmonary dysplasia: when inflammation meets organ development. *Mol Cell Pediatr* **3**, 23 (2016).

36. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biology* **18**, 83 (2017).

37. Bhandari, V. *et al.* Familial and Genetic Susceptibility to Major Neonatal Morbidities in Preterm Twins. *Pediatrics* **117**, 1901–1906 (2006).

38. Lavoie, P. M., Pham, C. & Jang, K. L. Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. *Pediatrics* **122**, 479–485 (2008).

39. Parad, R. B. *et al.* Role of Genetic Susceptibility in the Development of Bronchopulmonary Dysplasia. *J Pediatr* **203**, 234-241.e2 (2018).

40. Yu, K.-H., Li, J., Snyder, M., Shaw, G. M. & O'Brodovich, H. M. The genetic predisposition to bronchopulmonary dysplasia. *Curr Opin Pediatr* **28**, 318–323 (2016).

41. Lal, C. V. & Ambalavanan, N. Genetic predisposition to bronchopulmonary dysplasia. *Seminars in Perinatology* **39**, 584–591 (2015).

42. Lavoie, P. M. & Rayment, J. H. Genetics of bronchopulmonary dysplasia: An update. *Seminars in Perinatology* **47**, 151811 (2023).

43. Wang, H. *et al.* A Genome-Wide Association Study (GWAS) for Bronchopulmonary Dysplasia. *Pediatrics* **132**, 290–297 (2013).

44. Shaw, G. M. & O'Brodovich, H. M. Progress in understanding the genetics of bronchopulmonary dysplasia. *Semin Perinatol* **37**, 85–93 (2013).

45. Spencer, C. C. A., Su, Z., Donnelly, P. & Marchini, J. Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLOS Genetics* **5**, e1000477 (2009).

46. Alice Hadchouel *et al.* Identification of SPOCK2 As a Susceptibility Gene for Bronchopulmonary Dysplasia. *American Journal of Respiratory and Critical Care Medicine* **184**, 1164–1170 (2011).

47. Ambalavanan, N. *et al.* Integrated Genomic Analyses in Bronchopulmonary Dysplasia. *The Journal of Pediatrics* **166**, 531-537.e13 (2015).

48. Dai, D. *et al.* Bronchopulmonary Dysplasia Predicted by Developing a Machine Learning Model of Genetic and Clinical Information. *Frontiers in Genetics* **12**, (2021).

49. Hamvas, A. *et al.* Exome sequencing identifies gene variants and networks associated with extreme respiratory outcomes following preterm birth. *BMC Genetics* **19**, 94 (2018).

50. Luo, X. *et al.* Identification of genetic susceptibility in preterm newborns with bronchopulmonary dysplasia by whole-exome sequencing: BIVM gene may play a role. *Eur J Pediatr* https://doi.org/10.1007/s00431-022-04779-z (2023) doi:10.1007/s00431-022-04779-z.

51. Qiu, S. *et al.* Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Sig Transduct Target Ther* **8**, 1–37 (2023).

52. Hanna, M. H. & Brophy, P. D. Metabolomics in Neonatology. in *Metabolomics - Fundamentals and Applications* (IntechOpen, 2016). doi:10.5772/66295.

53. Piersigilli, F. & Bhandari, V. Metabolomics of bronchopulmonary dysplasia. *Clin Chim Acta* **500**, 109–114 (2020).

54. Atzori, L. *et al.* 1H NMR-based metabolomic analysis of urine from preterm and term neonates. *Front. Biosci. (Elite Ed)* **3**, 1005–1012 (2011).

55. Abdolahpour, S., Gholami, M., Mohsenipour, R. & Abbasi, F. Metabolic subtypes and biomarkers in preterm and term neonates via targeted screening. *Sci Rep* https://doi.org/10.1038/s41598-026-50955-8 (2026) doi:10.1038/s41598-026-50955-8.

56. Wilson, K. *et al.* Metabolomics of prematurity: analysis of patterns of amino acids, enzymes, and endocrine markers by categories of gestational age. *Pediatr Res* **75**, 367–373 (2014).

57. Yue, L., Lu, X., Dennery, P. A. & Yao, H. Metabolic dysregulation in bronchopulmonary dysplasia: Implications for identification of biomarkers and therapeutic approaches. *Redox Biol* **48**, 102104 (2021).

58. Kurzner, S. I. *et al.* Growth Failure in Infants With Bronchopulmonary Dysplasia: Nutrition and Elevated Resting Metabolic Expenditure. *Pediatrics* **81**, 379–384 (1988).

59. Ye, C. *et al.* Progressive Metabolic Abnormalities Associated with the Development of Neonatal Bronchopulmonary Dysplasia. *Nutrients* **14**, 3547 (2022).

60. Wang, L., Zhong, W. H., Liu, D. Y., Shen, H. Q. & He, Z. J. Metabolic analysis of infants with bronchopulmonary dysplasia under early nutrition therapy: An observational cohort study. *Exp Biol Med (Maywood)* **247**, 470–479 (2022).

61. Guo, Y. *et al.* Analysis of variable metabolites in preterm infants with bronchopulmonary dysplasia: a systematic review and meta-analysis. *Ital J Pediatr* **50**, 246 (2024).

62. Piersigilli, F. *et al.* Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics* **15**, 20 (2019).

63. La Frano, M. R. *et al.* Umbilical cord blood metabolomics reveal distinct signatures of dyslipidemia prior to bronchopulmonary dysplasia and pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L870–L881 (2018).

64. Frazer, L. C., Yakah, W. & Martin, C. R. Decreased Acetic Acid in the Stool of Preterm Infants Is Associated with an Increased Risk of Bronchopulmonary Dysplasia. *Nutrients* **14**, (2022).

65. Fanos, V. *et al.* Urinary metabolomics of bronchopulmonary dysplasia (BPD): preliminary data at birth suggest it is a congenital disease. *The Journal of Maternal-Fetal & Neonatal Medicine* **27**, 39–45 (2014).

66. Baraldi, E. *et al.* Untargeted Metabolomic Analysis of Amniotic Fluid in the Prediction of Preterm Delivery and Bronchopulmonary Dysplasia. *PLOS ONE* **11**, e0164211 (2016).

67. Pintus, M. C. *et al.* Urinary 1H-NMR Metabolomics in the First Week of Life Can Anticipate BPD Diagnosis. *Oxid Med Cell Longev* **2018**, 7620671 (2018).

68. Lal, C. V. *et al.* Early airway microbial metagenomic and metabolomic signatures are associated with development of severe bronchopulmonary dysplasia. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L810–L815 (2018).

69. You, Y. *et al.* Early metabolic markers as predictors of respiratory complications in preterm infants with bronchopulmonary dysplasia. *Early Human Development* **190**, 105950 (2024).

70. Bonadies, L. *et al.* Early prediction of bronchopulmonary dysplasia by urinary metabolomics: a case–control study. https://doi.org/10.1136/thorax-2025-223090 (2026) doi:10.1136/thorax-2025-223090.

71. Course, C. W. *et al.* Evidence of abnormality in glutathione metabolism in the airways of preterm born children with a history of bronchopulmonary dysplasia. *Sci Rep* **13**, 19465 (2023).

72. Chiu, C.-Y. *et al.* Metabolomic Analysis Reveals the Association of Severe Bronchopulmonary Dysplasia with Gut Microbiota and Oxidative Response in Extremely Preterm Infants. *Metabolites* **14**, 219 (2024).

73. Guo, Y. *et al.* Analysis of variable metabolites in preterm infants with bronchopulmonary dysplasia: a systematic review and meta-analysis. *Ital J Pediatr* **50**, 246 (2024).

74. Kastenmüller, G., Raffler, J., Gieger, C. & Suhre, K. Genetics of human metabolism: an update. *Hum Mol Genet* **24**, R93–R101 (2015).

75. Wang, N. *et al.* Genetic architecture and analysis practices of circulating metabolites in the NHLBI Trans-Omics for Precision Medicine Program. *The American Journal of Human Genetics* **112**, 2720–2738 (2025).

76. Hagenbeek, F. A. *et al.* Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun* **11**, 39 (2020).

77. Johnson, R. K. *et al.* Discovering metabolite quantitative trait loci in asthma using an isolated population. *Journal of Allergy and Clinical Immunology* **149**, 1807-1811.e16 (2022).

78. Gillenwater, L. A. *et al.* Plasma Metabolomic Signatures of Chronic Obstructive Pulmonary Disease and the Impact of Genetic Variants on Phenotype-Driven Modules. *Netw Syst Med* **3**, 159–181 (2020).

79. Mo, S. *et al.* Causal relationship between fatty acid metabolism and allergic asthma: a mendelian randomization and multi-omics study. *Journal of Asthma* **63**, 103–116 (2026).

80. Zhang, Y. *et al.* The relationship between genetic prediction of 486 blood metabolites and the risk of COPD: mendelian randomization study. *Sci Rep* **15**, 7349 (2025).

81. Ju, D., Hui, D., Hammond, D. A., Wonkam, A. & Tishkoff, S. A. Importance of Including Non-European Populations in Large Human Genetic Studies to Enhance Precision Medicine. *Annu Rev Biomed Data Sci* **5**, 321–339 (2022).

82. Haas, R. *et al.* Advancing Precision Health Discovery in a Genetically Diverse Health System. 2025.06.11.25329386 Preprint at https://doi.org/10.1101/2025.06.11.25329386 (2026).

83. Majara, L. *et al.* Low and differential polygenic score generalizability among African populations due largely to genetic diversity. *HGG Adv* **4**, 100184 (2023).

84. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250 (2022).

85. Yang, G., Mishra, M. & Perera, M. A. Multi-Omics Studies in Historically Excluded Populations: The Road to Equity. *Clin Pharmacol Ther* **113**, 541–556 (2023).

86. Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).

87. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).

88. Lu, C., Ahmed, R., Lamri, A. & Anand, S. S. Use of race, ethnicity, and ancestry data in health research. *PLOS Global Public Health* **2**, e0001060 (2022).

89. Ali-Khan, S. E., Krakowski, T., Tahir, R. & Daar, A. S. The use of race, ethnicity and ancestry in human genetic research. *Hugo J* **5**, 47–63 (2011).

90. Khan, A. T. *et al.* Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. *Cell Genom* **2**, 100155 (2022).

91. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research *et al. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. 26902 (National Academies Press, Washington, D.C., 2023). doi:10.17226/26902.

92. Lewis, A. C. F. *et al.* Getting genetic ancestry right for science and society. *Science* **376**, 250–252 (2022).

93. Liu, Y. *et al.* Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* **7**, 1 (2013).

94. Lewanski, A. L., Grundler, M. C. & Bradburd, G. S. The era of the ARG: An introduction to ancestral recombination graphs and their significance in empirical evolutionary genomics. *PLOS Genetics* **20**, e1011110 (2024).

95. Brandt, D. Y. C., Huber, C. D., Chiang, C. W. K. & Ortega-Del Vecchyo, D. The Promise of Inferring the Past Using the Ancestral Recombination Graph. *Genome Biol Evol* **16**, evae005 (2024).

96. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* **11**, 459–463 (2010).

97. Gopalan, S. *et al.* Human genetic admixture through the lens of population genomics. *Philos Trans R Soc Lond B Biol Sci* **377**, 20200410.

98. Spear, M. L. *et al.* Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. *Elife* **9**, e56029 (2020).

99. Reynolds, K. M. *et al.* Ancestry-driven metabolite variation provides insights into disease states in admixed populations. *Genome Med* **15**, 52 (2023).

100. Randolph, H. E., Aracena, K. A., Lin, Y.-L., Mu, Z. & Barreiro, L. B. Shaping immunity: The influence of natural selection on population immune diversity. *Immunological Reviews* **323**, 227–240 (2024).

101. Venkatesh, R. *et al.* Importance of genetic ancestry in pharmacogenomics for precision medicine. *Pharmacogenomics* **26**, 747–762.

102. Das, R., Tatarinova, T. V., Galieva, E. R. & Orlov, Y. L. Editorial: Association Between Individuals' Genomic Ancestry and Variation in Disease Susceptibility. *Front Genet* **13**, 831320 (2022).

103. Torgerson, D. G. *et al.* Ancestry and genetic associations with bronchopulmonary dysplasia in preterm infants. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L858–L869 (2018).

104. Qiang, Y.-X. *et al.* Genetic architecture of plasma metabolome in 254,825 individuals. *Nat Commun* **16**, 8272 (2025).

Chapter 2: Metabolome-wide Association Study of Bronchopulmonary Dysplasia in Extremely Premature Infants

Abstract

Rationale: Untargeted metabolomics may identify novel pathways involved in bronchopulmonary dysplasia (BPD), a poorly understood lung disease in extremely premature infants.

Objective: To identify metabolites and pathways altered in infants who develop BPD through a metabolome-wide association study.

Methods: We performed untargeted metabolomics on urine from 314 infants (<29 weeks gestational age) collected at two time points between postnatal days 5-15 and 23-32 with detailed time-series for 28 infants. We compared metabolite levels between infants with and without BPD at 36 weeks gestational age using logistic regression and performed a pathway enrichment analysis. A stratified analysis by nutrition type was performed at 23-32 days.

Measurements and Main Results: Among the 913 analyzed urinary metabolites, 23 and 68 were associated with BPD between postnatal days 5-15 and 23-32, respectively (p<0.05). We found enrichment in the lactoyl amino acid sub-pathway (5-15 days; p=0.005, FDR=0.39) and xanthine metabolism and fatty acid metabolism (23-32 days; p<0.002, FDR=0.04). Most BPD-differential metabolites were lower in infants with BPD at days 23-32 (OR=7.1, $p=2\times10^{-6}$), including infants receiving parenteral (OR=9.2, $p=7\times10^{-7}$) and full enteral nutrition (OR=6.1, p=0.003). Twenty-four metabolites, including vitamins C, B_1, and B_6, were lower in infants on parenteral nutrition with BPD at postnatal days 23-32.

Conclusion: BPD in extremely premature infants was associated with vitamin deficiencies and dysregulation of lactoyl amino acids, xanthine, and fatty acid metabolism. BPD-differential metabolites are lower in infants with BPD during the fourth week of life, potentially related to higher energy demand, metabolic dysregulation, nutritional deficiency, and/or impaired intestinal absorption.

Introduction

Bronchopulmonary dysplasia (BPD) is a developmental lung disease and major contributor to neurodevelopmental and growth abnormalities, long-term respiratory morbidity, and mortality among extremely low gestational age newborns.[1,2] Heterogeneity of BPD and advances in neonatal healthcare have led to evolving definitions.[3] The recently described "new BPD" in extremely premature infants is characterized by impaired alveolar development and dysregulated pulmonary vasculature, leading to persistent abnormalities in gas exchange, airway function, respiratory mechanics, and lung volumes.[4] Risk factors associated with BPD include low gestational age and birthweight, male sex, and social and environmental factors that sometimes correlate with race/ethnicity.[5,6] Despite advances in neonatal healthcare, there has been minimal change in the incidence of BPD during the last 20 years.[7] The limited understanding of BPD pathophysiology has hindered the development of effective diagnostics, preventative therapies, and treatments that may help lower the incidence of BPD.[7, 8] Novel approaches are needed to identify the biological mechanisms involved in BPD.

Metabolomics involves the quantification of many hundreds of biochemicals (metabolites) that are present in a biological sample and has emerged as a powerful approach for studies of complex disease. It can simultaneously capture the combined effects of environmental and genetic factors that may impact human physiology, and increase our overall comprehension of disease pathogenesis.[9] Previous studies have implicated metabolic dysregulation in amino acid, lipid, and/or carnitine pathways in BPD; however, there is a limited consensus across studies.[10--19] This may be due to varying infant demographics, small sample sizes of individual studies (<100 preterm infants), differences in the type of biofluid sampled, and reliance on single time point sampling during the clinical course of premature infants in the neonatal intensive care unit (NICU). Furthermore, the majority of studies have used a targeted approach to characterize metabolites

of interest, whereas untargeted metabolomics can query a greater number of metabolites to test for associations with disease and provide a broader view of the pathways involved.

In this study, we characterized the urinary metabolome of 314 extremely premature infants at high risk of developing BPD (including 178 infants diagnosed with BPD at 36 weeks) using an untargeted approach to identify a broad array of metabolites present in urine. We performed a metabolome-wide association study (MWAS) for BPD at 36 weeks gestational age to identify metabolites and metabolic pathways associated with BPD, including multiple time point assessments in the first 30 days of life. Our results provide insight into the metabolic pathways involved in BPD development and offer information to support early diagnosis and the development of more effective therapies and interventions. Some of the results of this study have been previously reported in abstract form.[20]

Methods

Study Description

A total of 171 infants from the Trial of Late Surfactant for Prevention of Bronchopulmonary Dysplasia (TOLSURF, NCT01022580) and 143 infants from the Prematurity and Respiratory Outcomes Program (PROP, NCT01435187) were included in our study (details in the extended methods of the supplementary material). Clinical data were obtained for both cohorts and summarized in Table 2.1. BPD was assessed at 36 weeks post-menstrual age by physiological testing.[21] Research protocols were approved by the institutional review boards of the participating institutions, and a parent of each infant provided written informed consent.

Table 2.1. Characteristics of infants from TOLSURF and PROP included in metabolomic studies by BPD status at 36 weeks postmenstrual age.

	TOLSURF (n = 171)			PROP (n = 143)		
	BPD	**No BPD**	***p***	**BPD**	**No BPD**	***p***
Number of Infants	106	65	-	72	71	-
Number of Infants by Maternal Race†			0.011			1
Black or African American	30 (28)	31 (48)		30 (42)	30 (42)	
Hispanic/Latine	14 (13)	11 (17)		12 (16)	11 (16)	
Non-Hispanic White	62 (59)	23 (35)		30 (42)	30 (42)	
Gestational Age in weeks†	25.3 (1.3)	25.4 (1.1)	0.58	25.9 (1.1)	25.8 (1.3)	0.42
Sex (male)†	64 (60.4)	36 (55.4)	0.53	41 (56.9)	34 (47.9)	0.32
Birthweight (g)†	706 (172)	757 (177)	0.064	766 (135)	805 (163)	0.12
Average RSS days 7-14	4.5 (2.1)	3.5 (1.3)	$1.2x10^{-4}$	2.8 (1.7)	2.4 (1.6)	0.077
Average RSS days 21-28	4.4 (2.3)	3.7 (2.1)	$1.8x10^{-15}$	4.1 (3.0)	1.7 (1.3)	$2.8x10^{-15}$
Age days sample 1	9.4 (2.4)	9.8 (2.3)	0.35	6.6 (1.1)	6.8 (1.0)	0.40
Age days sample 2	27.5 (2.1)	27.7 (2.3)	0.68	28.0 (1.0)	28.3 (0.86)	0.13
BPD at 40 weeks	67 (65.1)	65 (0)	-	61.8	71 (0)	-
On PN sample 1	106 (100)	65 (100)	1	72 (100)	71 (100)	1
On PN sample 2*	67 (63.2)	42 (64.6)	0.87	44 (61.1)	51 (71.8)	0.22
On corticosteroids sample 1†	8 (7.5)	6 (9.2)	0.78	3 (4.2)	1 (1.4)	0.62
On corticosteroids sample 2	22 (20.8)	14 (21.5)	1	20 (27.8)	21 (29.6)	0.85
On caffeine sample 1	83 (78.3)	55 (84.6)	0.42	51 (70.8)	53 (74.6)	0.71
On caffeine sample 2	90 (84.9)	60 (92.3)	0.22	60 (83.3)	65 (91.5)	0.21

TOLSURF Trial of Late Surfactant for Prevention of Bronchopulmonary Dysplasia; *PROP* Prematurity and Respiratory Outcomes Program; *BPD* Bronchopulmonary Dysplasia; *p* p-value; *RSS* Respiratory Severity Score; *PN* Parenteral Nutrition

Values are represented by the mean (standard error) for continuous variables and the count (percent) for categorical variables. Quantitative measurements were compared using a t-test, and categorical data were compared using a Fisher's Exact test.

†Covariates included in the MWAS.
*Additional covariate included in the MWAS at time point 2.

Sample Selection

Urine samples were collected from each infant and stored at -80°C until processing. Infants were selected based on survival at 40 weeks gestational age and availability of urine samples. Two samples were selected for each infant, one collected between postnatal days 5-15 (time point 1, mean/SD 8±2 days) and another between postnatal days 23-32 (time point 2, mean/SD 28±2 days).

Metabolomics Profiling

Untargeted metabolomic assays were performed using Ultra High-Performance Liquid Chromatography-Tandem Mass Spectrometry (UHPLC/MS-MS) by Metabolon Inc. as previously described.[22, 23] Metabolic profiling was conducted separately for each cohort, with anchor samples included to facilitate the comparison of metabolite abundance between cohorts. Metabolite levels were normalized to osmolality and median scaled. Metabolites with >30% of values below the limit of detection within each cohort were removed from further analysis. A detailed description of metabolomic profiling is provided in the extended methods in the supplementary material.

Statistical Analyses

Principal component analysis (PCA) was performed across all metabolites to evaluate global differences in the urinary metabolome. A Mann-Whitney *U* test was performed to assess differences in the first two principal components (PCs) for infants with and without BPD, including sub-analyses at time point 2 for infants receiving parenteral vs. full enteral nutrition to investigate potential confounding.

A metabolome-wide association study (MWAS) was performed using logistic regression to test for associations between BPD and quantitative levels of individual metabolites at each time point and the fold-change in levels at time point 2 vs. time point 1. Covariates included clinical and demographic characteristics (Table 2.1). Analyses were done within each cohort and combined using a fixed-effects meta-analysis. Sub-analyses were performed by nutrition type at time point 2 to further evaluate nutrition source as a potential confounder in associations with BPD. Details on infant feeding status categorization, statistical models, and covariates are included in the extended methods of the supplementary material.

Metabolites were defined as “differential” when associated with BPD at a threshold of $p<0.05$. Metabolites of known identity were annotated to 88 and 91 sub-pathways at time points 1 and 2, respectively (Metabolon Inc, April 2025). Pathway enrichment analysis was performed for each super-pathway and sub-pathway using a Fisher’s Exact Test for each MWAS comparison (additional details in extended methods of the supplementary material).

Longitudinal data for selected differential metabolites from a prior study[23] were aligned to days relative to discontinuation of parenteral nutrition and visualized using time course plots with locally weighted scatterplot smoothing by BPD status. Data included untargeted metabolomics of 302 urine samples from 28 infants (17 with BPD and 11 without BPD) in TOLSURF (additional details in extended methods of the supplementary material), of which 25 were also included in the current two-time point study.

Analyses were performed using Python 3.11.4[24] and R 4.5.1.[25]

Results

Clinical and demographic characteristics were similar between infants with and without BPD at 36 weeks gestational age, apart from infants with BPD having higher respiratory severity scores (RSS) at both time points (Table 2.1). Mean values across groups for gestational age ranged between 25-26 weeks, and birth weight between 720-790 g. Untargeted urinary metabolomic profiling identified 1441 unique metabolites across both cohorts at time points 1 and 2, of which 874 and 913 passed quality control thresholds, respectively (851 overlapping between time points). The first two principal components across all metabolites showed no separation of infants by BPD status ($p>0.05$; Figure 2.1A-D; Figure S2.1A-B, S2.1D), apart from infants on parenteral nutrition at time point 2 (PC1; $p=0.042$; Figure S2.1C).

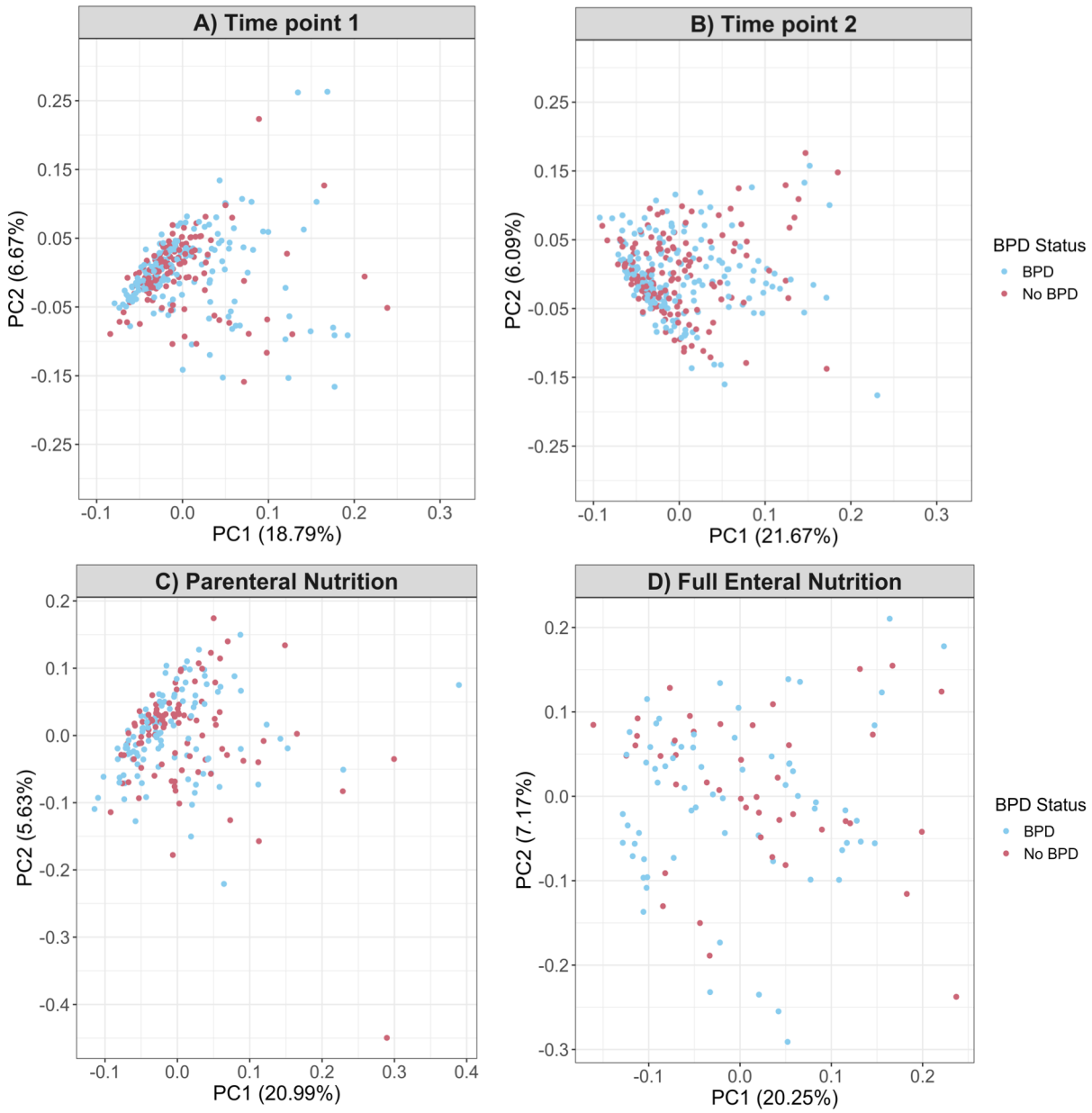

Figure 2.1. Principal component analyses (PCA) of urinary metabolite levels of infants from both cohorts at two sampling time points.

Scatter plots of the first two principal components (PC1 and PC2). (A) PCA of 874 biochemicals detected in urine samples collected between postnatal days 5-15 (time point 1) across all infants. (B) PCA of 913 biochemicals detected in urine samples collected between postnatal days 23-32 (time point 2) across all infants. (C) PCA of the same 913 biochemicals at time point 2 among infants on parenteral nutrition, and (D) among infants on full enteral nutrition. Percent values on each axis indicate the variance explained.

Metabolome-Wide Association Study (MWAS)

We tested for an association between BPD and levels of urinary metabolites for 1) each time point separately, 2) fold-change across time points, and 3) sub-analyses at time point 2 with infants separated by feeding modality (parenteral vs. full enteral).

Levels of 23 and 68 individual metabolites were associated with BPD at $p<0.05$ at time points 1 and 2, respectively (Figure 2.2A-B and Supplemental Table S2.1-S2.2), and 28 metabolites were associated with fold-change (Table S2.3). Odds ratios for top metabolites associated with BPD are shown in Figure 2.3, with cohort-specific effects in the same direction. Across BPD-differential metabolites ($p<0.05$), 13 and 44 were of known identity at time points 1 and 2, respectively, and belonged to a variety of sub-pathways (Table 2.2, Table 2.3). Five metabolites were associated with BPD at $p<0.05$ at both time points, including 3-hydroxykynurenine (tryptophan metabolism), pipecolate (lysine metabolism), beta-hydroxyisovaleroylcarnitine (leucine/isoleucine/valine metabolism), and two unnamed metabolites. One of these, 3-hydroxykynurenine, showed an opposite direction of association with BPD across time points, with higher levels in infants with BPD at time point 1 and lower levels at time point 2. A similar trend was observed in longitudinal data following discontinuation of parenteral nutrition (Figure 2.4A). Pipecolate was higher in infants with BPD at both time points, and beta-hydroxyisovaleroylcarnitine and two unnamed metabolites were lower in infants with BPD at both time points. Fold-changes over time in the levels of 19 metabolites of known identity were associated with BPD at $p<0.05$ among all infants, including hexanoylcarnitine (C6) and four unnamed metabolites that were also associated with BPD at $p<0.05$ at time point 2.

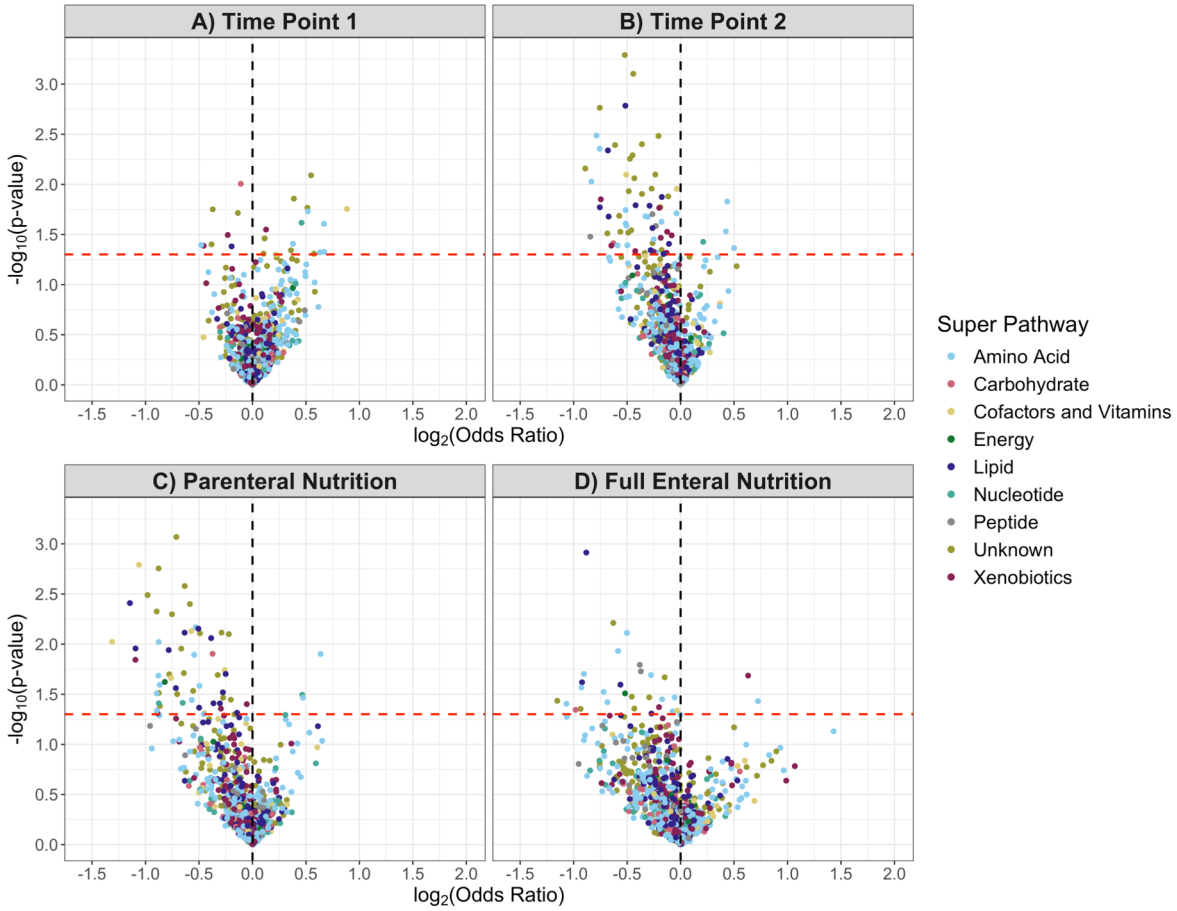

Figure 2.2. Volcano plot identifying differential metabolites that vary by BPD status at two sampling time points.

Metabolite odds ratios (OR) are presented as $\log_2$(OR), and the significance level of difference as -log10(*p*-value). The dashed red line represents $p<0.05$, and the dashed black line divides metabolites higher in infants who developed BPD ($\log_2$(OR)>0) from those higher in infants who did not ($\log_2$(OR)<0). Each metabolite is represented by a dot colored according to its super pathway. (A) Days 5-15 postnatal (time point 1) across all infants, (B) Days 23-32 postnatal (time point 2) across all infants, (C) Time point 2 among infants on parenteral nutrition, (D) Time point 2 among infants on full enteral nutrition.

Sub-analyses among infants separated by mode of nutrition at time point 2 identified 60 BPD-differential metabolites in those receiving parenteral nutrition and 30 in those receiving enteral nutrition (Figure 2.2C-D; Supplemental Tables S2.4-S2.5), of which three overlapped in the same direction of effect [androstenediol (3beta,17beta) disulfate (2), 5-(galactosylhydroxy)-lysine, and an unnamed metabolite]. A total of 28 were unique to infants receiving parenteral nutrition (Table 2.4), and 13 were unique to infants on full enteral nutrition (Table 2.5). Of the 68

metabolites implicated in BPD at time point 2 across all infants, 29 were BPD-differential in sub-analyses of infants on parenteral nutrition, and 14 in infants on enteral nutrition (Figure 2.5).

Table 2.2. Differential metabolites (p<0.05) of known identity during postnatal days 5-15 (time point 1).

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
sucrose	Disaccharides and Oligosaccharides	0.074 (0.156)	0.638	-0.081 (0.030)	0.007	-0.076 (0.029)	0.010
4-methylnonanoylcarnitine	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.021 (0.104)	0.838	-0.244 (0.087)	0.005	-0.136 (0.067)	0.042
methyl vanillate sulfate	Food Component/Plant	0.137 (0.074)	0.064	0.067 (0.048)	0.155	0.088 (0.040)	0.028
menthol glucuronide		-0.079 (0.191)	0.680	-0.175 (0.081)	0.031	-0.160 (0.075)	0.032
5-aminolevulinate	Hemoglobin and Porphyrin Metabolism	0.526 (0.326)	0.106	0.756 (0.422)	0.073	0.612 (0.258)	0.018
N-lactoyl isoleucine	Lactoyl Amino Acid	0.389 (0.265)	0.143	0.546 (0.398)	0.171	0.437 (0.221)	0.048
N-lactoyl leucine		0.469 (0.290)	0.106	0.458 (0.398)	0.250	0.465 (0.234)	0.047
beta-hydroxyisovaleroylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.093 (0.248)	0.707	-0.512 (0.214)	0.017	-0.332 (0.162)	0.040
pipecolate	Lysine Metabolism	0.405 (0.247)	0.100	0.599 (0.377)	0.112	0.463 (0.206)	0.025
guanine	Purine Metabolism, Guanine containing	0.144 (0.161)	0.371	0.904 (0.295)	0.002	0.318 (0.141)	0.024
3-hydroxykynurenine	Tryptophan Metabolism	0.349 (0.175)	0.045	0.388 (0.312)	0.214	0.359 (0.152)	0.019
tyrosine	Tyrosine Metabolism	0.163 (0.133)	0.222	1.253 (0.421)	0.003	0.262 (0.127)	0.039
caffeine	Xanthine Metabolism	-0.385 (0.205)	0.060	-0.226 (0.237)	0.342	-0.317 (0.155)	0.041

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Analysis used logistic regression for the TOLSURF and PROP cohorts separately, adjusting for birth weight, sex, self-reported maternal race/ethnicity, and corticosteroids (>1 day), followed by a meta-analysis. A negative effect size/unit change (beta) indicates metabolites of lower abundance in BPD infants. Metabolites are grouped by sub-pathway. Sub-pathways and metabolites are listed alphabetically. Complete results are provided in the supplementary material.

Table 2.3. Differential metabolites (p<0.05) of known identity during postnatal days 23-32 (time point 2).

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
N-acetylglucosaminylasparagine	Aminosugar Metabolism	-0.470 (0.271)	0.084	-0.385 (0.335)	0.251	-0.436 (0.211)	0.039
androstenediol (3beta,17beta) disulfate (1)	Androgenic Steroids	-0.161 (0.134)	0.227	-0.231 (0.141)	0.102	-0.194 (0.097)	0.045
androstenediol (3beta,17beta) disulfate (2)		-0.557 (0.239)	0.02	-0.391 (0.231)	0.091	-0.471 (0.166)	0.005
ascorbate (vitamin C)	Ascorbate and Aldarate Metabolism	-0.025 (0.010)	0.013	-0.032 (0.056)	0.571	-0.025 (0.010)	0.011
diaminopimelate	Bacterial/Fungal	-0.194 (0.130)	0.134	-0.159 (0.120)	0.186	-0.175 (0.088)	0.047
sulfate*	Chemical	-0.533 (0.301)	0.077	-0.354 (0.321)	0.271	-0.449 (0.220)	0.041
3-hydroxypyridine sulfate		-0.113 (0.071)	0.113	-0.168 (0.115)	0.143	-0.128 (0.061)	0.034
phenylalanylhydroxyproline*	Dipeptide	-0.286 (0.111)	0.01	-0.079 (0.111)	0.479	-0.183 (0.078)	0.02
leucylhydroxyproline*	Dipeptide Derivative	-0.205 (0.092)	0.026	-0.089 (0.116)	0.443	-0.160 (0.072)	0.026
pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	-0.206 (0.120)	0.086	-0.197 (0.117)	0.093	-0.201 (0.084)	0.016
3-hydroxyoctanoylcarnitine (1)	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.053 (0.101)	0.604	-0.440 (0.143)	0.002	-0.182 (0.083)	0.027
hexanoylcarnitine (C6)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.079 (0.058)	0.173	-0.236 (0.094)	0.012	-0.122 (0.049)	0.013
octanoylcarnitine (C8)		-0.054 (0.037)	0.143	-0.861 (0.226)	$1x10^{-4}$	-0.075 (0.037)	0.039
2-methylhexanoylcarnitine		-0.056 (0.061)	0.362	-0.428 (0.147)	0.004	-0.111 (0.056)	0.049
2-methylmalonylcarnitine (C4-DC)	Fatty Acid Metabolism (also BCAA Metabolism)	-0.280 (0.242)	0.247	-0.902 (0.369)	0.014	-0.467 (0.202)	0.021
2-hydroxyadipate	Fatty Acid, Dicarboxylate	-0.447 (0.277)	0.107	-0.653 (0.359)	0.069	-0.524 (0.219)	0.017
erythritol	Food Component/ Plant	-0.595 (0.301)	0.048	-0.442 (0.295)	0.134	-0.517 (0.211)	0.014
1-methylguanidine	Guanidino and Acetamido Metabolism	-0.128 (0.120)	0.285	-0.439 (0.174)	0.012	-0.228 (0.099)	0.021

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
4-imidazoleacetate	Histidine Metabolism	0.426 (0.265)	0.108	0.289 (0.224)	0.197	0.346 (0.171)	0.043
1-ribosyl-imidazoleacetate*		0.374 (0.192)	0.052	0.251 (0.162)	0.122	0.302 (0.124)	0.015
N-acetyl-1-methylhistidine*		-0.351 (0.182)	0.053	-0.369 (0.271)	0.174	-0.357 (0.151)	0.018
beta-hydroxyisovaleroylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.267 (0.225)	0.236	-1.053 (0.322)	0.001	-0.525 (0.185)	0.004
tiglylcarnitine (C5:1-DC)		-0.135 (0.184)	0.463	-0.980 (0.310)	0.002	-0.356 (0.158)	0.025
3-methylglutarylcarnitine (2)		-0.078 (0.123)	0.523	-0.389 (0.140)	0.006	-0.213 (0.092)	0.021
methylsuccinoylcarnitine		-0.095 (0.082)	0.246	-0.147 (0.073)	0.045	-0.124 (0.055)	0.024
pipecolate	Lysine Metabolism	0.395 (0.193)	0.041	0.198 (0.188)	0.293	0.294 (0.135)	0.029
5-(galactosylhydroxy)-lysine		-0.619 (0.243)	0.011	-0.442 (0.286)	0.123	-0.545 (0.185)	0.003
hydroxy-N6,N6,N6-trimethyllysine*		-0.593 (0.285)	0.038	-0.555 (0.356)	0.119	-0.578 (0.223)	0.009
taurine	Methionine, Cysteine, SAM and Taurine Metabolism	-0.097 (0.114)	0.396	-0.209 (0.109)	0.054	-0.155 (0.079)	0.048
mevalonate	Mevalonate Metabolism	-0.385 (0.157)	0.014	-0.328 (0.165)	0.046	-0.358 (0.114)	0.002
N,N-dimethyl-pro-pro	Modified Peptides	-0.648 (0.374)	0.083	-0.511 (0.406)	0.208	-0.585 (0.275)	0.033
isoputreanine	Polyamine Metabolism	-0.308 (0.239)	0.197	-0.639 (0.362)	0.078	-0.408 (0.199)	0.041
21-hydroxypregnenolone disulfate	Pregnenolone Steroids	-0.326 (0.180)	0.07	-0.263 (0.164)	0.109	-0.292 (0.121)	0.016
uridine	Pyrimidine Metabolism, Uracil containing	0.101 (0.106)	0.34	0.182 (0.094)	0.053	0.146 (0.070)	0.037
riboflavin (vitamin B2)	Riboflavin Metabolism	-0.494 (0.194)	0.011	-0.227 (0.183)	0.216	-0.353 (0.133)	0.008
5-hydroxyindoleacetate	Tryptophan Metabolism	-0.544 (0.233)	0.02	-0.129 (0.243)	0.595	-0.345 (0.168)	0.040
3-hydroxykynurenine		-0.447 (0.198)	0.024	-0.116 (0.091)	0.202	-0.174 (0.083)	0.035
4-hydroxyphenylacetoylcarnitine	Tyrosine Metabolism	-0.024 (0.014)	0.081	-0.032 (0.020)	0.112	-0.026 (0.011)	0.019
pro-hydroxy-pro	Urea cycle; Arginine and Proline Metabolism	-0.660 (0.269)	0.014	-0.190 (0.194)	0.328	-0.351 (0.157)	0.026

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
3-methylxanthine	Xanthine Metabolism	-0.050 (0.043)	0.24	-0.063 (0.035)	0.071	-0.058 (0.027)	0.032
7-methylxanthine		-0.061 (0.057)	0.281	-0.145 (0.077)	0.061	-0.090 (0.046)	0.048
1,7-dimethylurate		-0.158 (0.090)	0.079	-0.124 (0.076)	0.102	-0.138 (0.058)	0.017
1-methylurate		-0.496 (0.241)	0.04	-0.173 (0.186)	0.354	-0.294 (0.147)	0.046
7-methylurate		-0.058 (0.052)	0.267	-0.129 (0.063)	0.039	-0.087 (0.040)	0.030

Meta meta-analysis; Beta beta-coefficient; SE standard error; p p-value

Vitamins are in boldface.

Analysis used logistic regression for the TOLSURF and PROP cohorts separately, adjusting for birth weight, sex, self-reported maternal race/ethnicity, nutrition source (parenteral vs enteral nutrition), and corticosteroids (>1 day), followed by a meta-analysis. A negative effect size/unit change (beta) indicates metabolites of lower abundance in BPD infants. Metabolites are grouped by sub-pathway. Sub-pathways and metabolites are listed alphabetically. Complete results are provided in the supplementary material.

Levels of metabolites at time point 2 were asymmetric in the direction of effect in BPD, including sub-analyses by mode of nutrition (Figure 2.2B, 2.2C, and 2.2D). Among all infants, 94% of differential metabolites (64 of 68) were lower in infants with BPD at time point 2 (OR=7.1, p=2.2x10^{-6}; Supplemental Table S2.2), with increased odds for sub-analyses of infants on parenteral nutrition (57 of 60 [95%], OR=9.2, p=7.2x10^{-7}; Supplemental Table S2.4) as compared to infants on full enteral nutrition (28 of 30 [93%], OR=6.1, p=0.003; Supplemental Table S2.5). This asymmetry was not observed at time point 1, where only 8 of 23 differential metabolites were at lower levels in infants with BPD (35%, OR=0.81, p=0.83). Longitudinal trends for select metabolites implicated through MWAS using a time-course design in a smaller number of infants demonstrate the trajectory of reduced metabolite levels after stopping parenteral nutrition for infants who develop BPD (Figure 2.4).

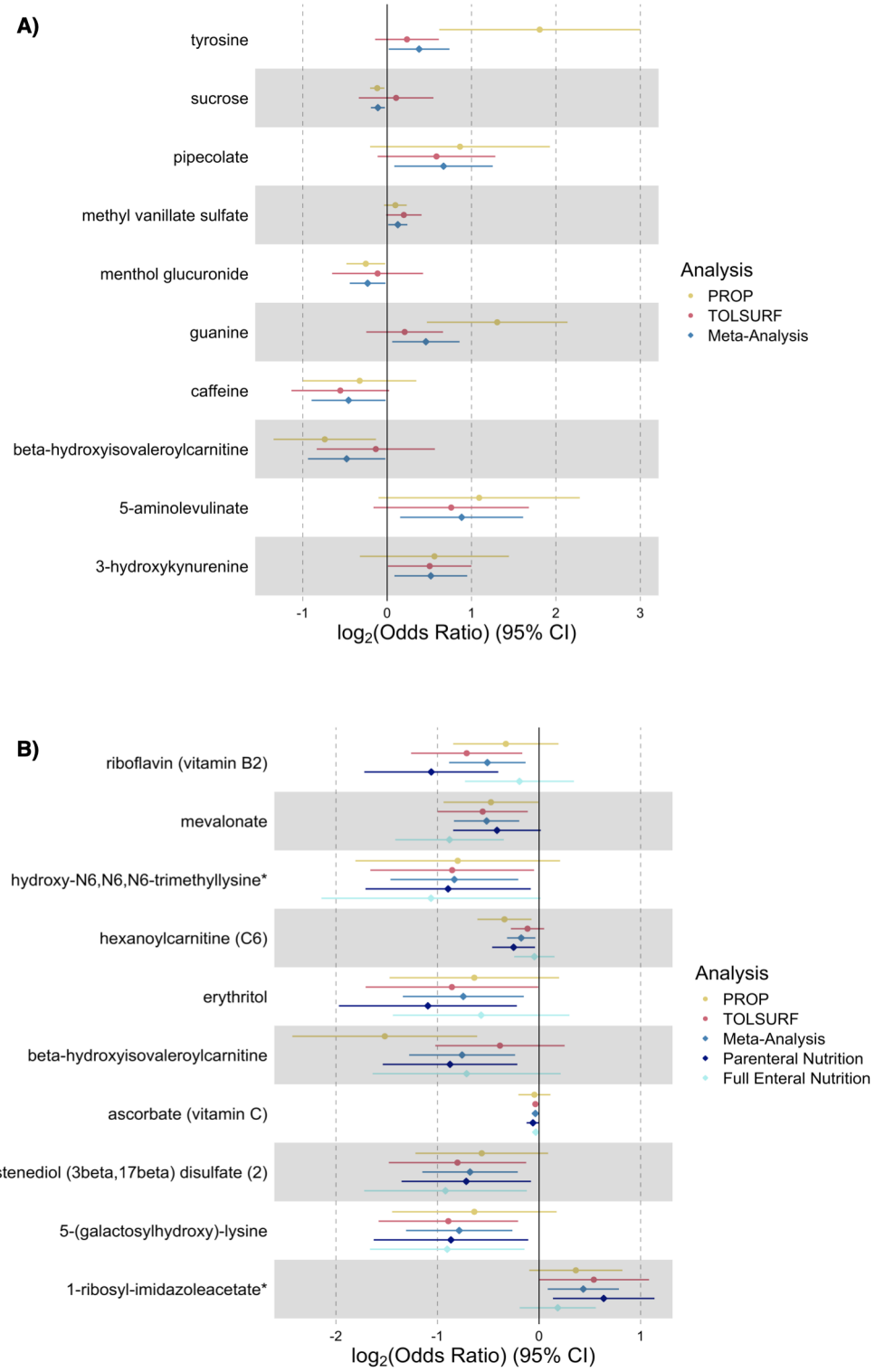

Figure 2.3. Odds ratios of the top 10 differential metabolites of known identity associated with BPD.

Metabolites are in ascending *p*-value order from top to bottom. Horizontal error bars denote 95% confidence intervals. The solid vertical black line divides metabolites higher in infants who developed BPD ($log_2(OR)>0$) from those higher in infants who did not ($log_2(OR)<0$). (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) Odds ratio estimates for metabolites in the analysis in infants from the PROP cohort are colored in yellow, from the TOLSURF cohort in red, and for the meta-analysis in blue for the top 10 differential metabolites at (A) time point 1 and (B) time point 2. Time point 2 also includes odds ratios from the feeding sub-analyses, represented by dark blue for infants on parenteral nutrition and light blue for those on full enteral nutrition.

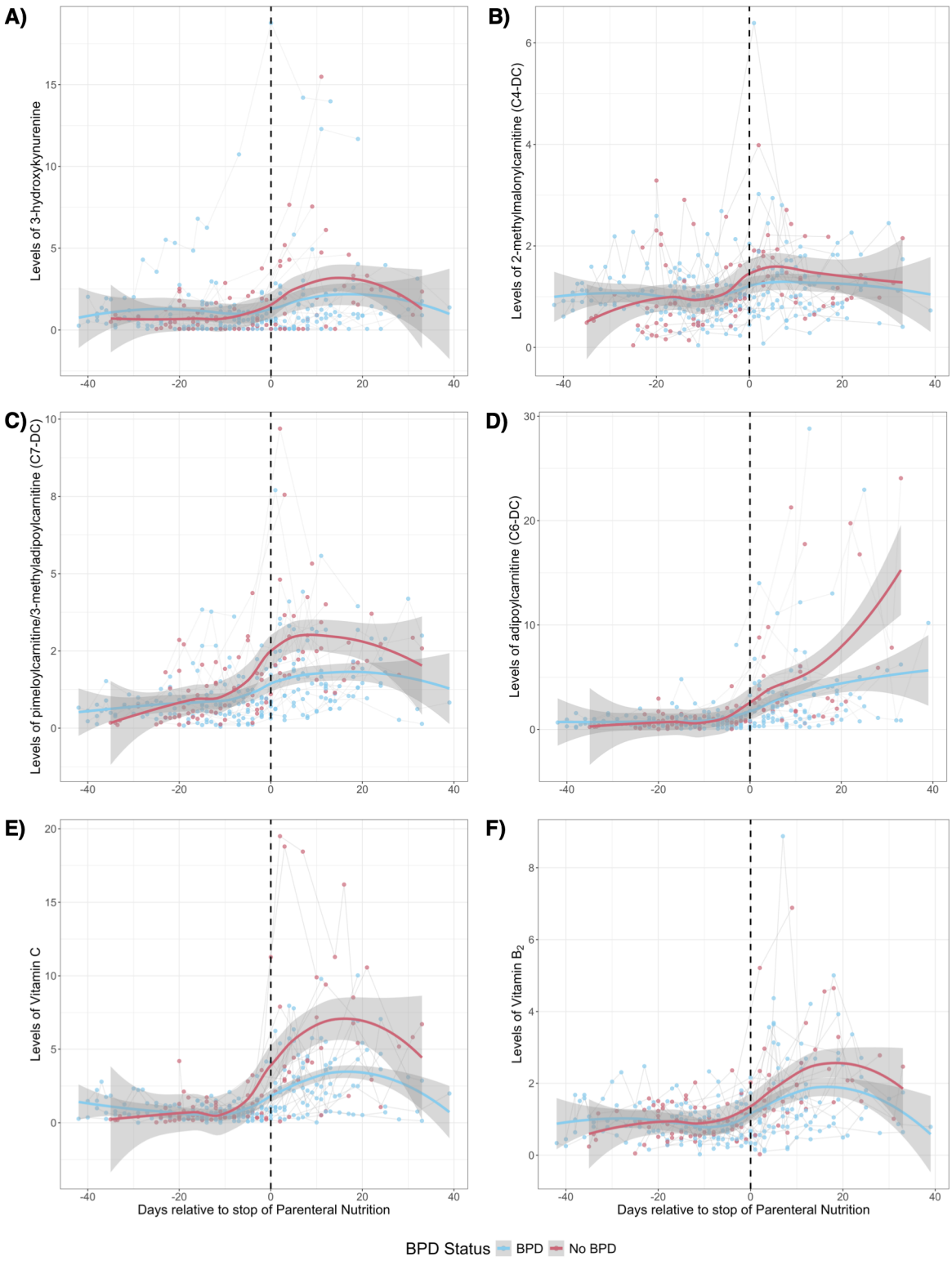

Figure 2.4. Longitudinal trajectories of selected individual metabolites relative to parenteral nutrition discontinuation by BPD status.

(Figure caption appears on the next page)

(Figure caption from previous page) Local estimated scatterplot smoothing (LOESS) plots are shown for (A) 3-hydroxykynurenine, (B) 2-methylmalonylcarnitine (C4-DC), (C) pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC), (D) adipoylcarnitine (C6-DC), (E) vitamin C (ascorbic acid 3-sulfate*), and (F) vitamin B_2 (riboflavin). Points represent individual observations, with connected points indicating measurements from the same infant. Shaded bands represent 95% confidence intervals around the LOESS fit. The dashed vertical line indicates the day parenteral nutrition was discontinued (day 0). For all metabolites shown, increasing levels are generally observed after stopping parenteral nutrition, with a slower increase in infants who develop BPD.

Table 2.4. Differential metabolites (p<0.05) of known identity unique to infants on parenteral nutrition during postnatal days 23-32 (time point 2).

Metabolite Name	Sub-pathway	All Infants		Parenteral Nutrition		Enteral Nutrition	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
ascorbic acid 3-sulfate*	Ascorbate and Aldarate Metabolism	-0.085 (0.046)	0.064	-0.268 (0.102)	0.007	-0.027 (0.043)	0.465
dehydroascorbate		-0.245 (0.165)	0.106	-0.608 (0.294)	0.018	-0.013 (0.255)	0.386
lactose	Disaccharides and Oligosaccharides	-0.053 (0.040)	0.093	-0.253 (0.123)	0.012	-0.020 (0.044)	0.858
adipoylcarnitine (C6-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	0.113 (0.075)	0.865	0.323 (0.150)	0.030	0.030 (0.063)	0.784
suberoylcarnitine (C8-DC)		-0.180 (0.145)	0.492	-0.610 (0.264)	0.048	0.051 (0.190)	0.914
(S)-3-hydroxybutyrylcarnitine	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.003 (0.017)	0.065	-0.192 (0.089)	0.009	0.006 (0.021)	0.531
3-hydroxyhexanoylcarnitine (1)		-0.134 (0.072)	0.179	-0.397 (0.148)	0.039	-0.065 (0.089)	0.645
hexanoylglycine	Fatty Acid Metabolism (Acyl Glycine)	-0.053 (0.033)	0.082	-0.181 (0.076)	0.039	-0.029 (0.034)	0.859
malonylcarnitine	Fatty Acid Synthesis	-0.234 (0.214)	0.067	-0.616 (0.313)	0.011	0.177 (0.336)	0.947
dimethylmalonic acid	Fatty Acid, Dicarboxylate	-0.065 (0.040)	0.275	-0.205 (0.102)	0.049	-0.025 (0.048)	0.599
furaneol sulfate	Food Component/Plant	-0.085 (0.049)	0.102	-0.211 (0.102)	0.044	0.012 (0.065)	0.607
piperidine		-0.115 (0.069)	0.564	-0.259 (0.104)	0.040	0.020 (0.111)	0.612
6-oxopiperidine-2-carboxylate	Lysine Metabolism	-0.281 (0.153)	0.216	-0.543 (0.215)	0.021	0.018 (0.267)	0.788
N6-carboxyethyllysine		-0.081 (0.060)	0.567	-0.316 (0.161)	0.013	-0.056 (0.070)	0.717
1-methylhypoxanthine	Purine Metabolism, (Hypo)Xanthine/Inosine containing	0.132 (0.107)	0.137	0.327 (0.154)	0.039	-0.158 (0.164)	0.958
uric acid ribonucleoside*		-0.456 (0.240)	0.114	-0.619 (0.310)	0.039	-0.336 (0.415)	0.886

Metabolite Name	Sub-pathway	All Infants		Parenteral Nutrition		Enteral Nutrition	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
5-methyluridine (ribothymidine)	Pyrimidine Metabolism, Uracil containing	-0.025 (0.044)	0.131	-0.378 (0.152)	0.032	0.019 (0.053)	0.636
succinylcarnitine (C4-DC)	TCA Cycle	-0.004 (0.006)	0.081	-0.037 (0.018)	0.024	0.004 (0.008)	0.714
thiamin (vitamin B1)	Thiamine Metabolism	-0.251 (0.221)	0.065	-0.910 (0.351)	0.022	0.211 (0.331)	0.918
methylurea	Urea cycle; Arginine and Proline Metabolism	-0.107 (0.064)	0.178	-0.312 (0.158)	0.050	-0.031 (0.089)	0.424
N-monomethylarginine		-0.008 (0.012)	0.214	-0.100 (0.051)	0.034	-0.001 (0.014)	0.336
N,N,N-trimethyl-alanylproline betaine (TMAP)		-0.312 (0.179)	0.058	-0.568 (0.251)	0.046	-0.114 (0.310)	0.418
pyridoxate	Vitamin B6 Metabolism	-0.335 (0.181)	0.256	-0.528 (0.230)	0.010	-0.036 (0.348)	0.524
pyridoxine (vitamin B6)		-0.284 (0.180)	0.097	-0.453 (0.219)	0.049	0.053 (0.369)	0.730

Beta beta-coefficient; *SE* standard error; *p* *p*-value

Vitamins are in **boldface**.

Analysis used logistic regression for the TOLSURF and PROP cohorts separately, stratified by nutrition source (parenteral vs enteral nutrition), adjusting for birth weight, sex, self-reported maternal race/ethnicity, and corticosteroids (>1 day), followed by a meta-analysis. A negative effect size/unit change (beta) indicates metabolites of lower abundance in BPD infants. Metabolites are grouped by sub-pathway. Sub-pathways and metabolites are listed alphabetically. Complete results are provided in the supplementary material.

Table 2.5. Differential metabolites (*p*<0.05) of known identity unique to infants on enteral nutrition during postnatal days 23-32 (time point 2).

Metabolite Name	Sub-pathway	All Infants		Parenteral Nutrition		Enteral Nutrition	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
diisopropanolamine	Chemical	0.009 (0.090)	0.921	-0.105 (0.119)	0.376	0.438 (0.189)	0.021
cysteinylglycine disulfide*	Glutathione Metabolism	-0.008 (0.090)	0.931	0.040 (0.130)	0.760	-0.643 (0.291)	0.027
2,3-dihydroxy-3-methylvalerate	Leucine, Isoleucine and Valine Metabolism	-0.006 (0.027)	0.831	0.039 (0.038)	0.304	-0.120 (0.057)	0.035
S-methylcysteine sulfoxide	Methionine, Cysteine, SAM and Taurine Metabolism	-0.087 (0.056)	0.119	-0.016 (0.078)	0.841	-0.192 (0.083)	0.020
4-hydroxyphenylacetate	Phenylalanine Metabolism	-0.036 (0.019)	0.062	-0.010 (0.037)	0.789	-0.053 (0.027)	0.047
5-methylthioadenosine (MTA)	Polyamine Metabolism	0.241 (0.125)	0.053	0.085 (0.167)	0.611	0.502 (0.241)	0.037

Metabolite Name	Sub-pathway	All Infants		Parenteral Nutrition		Enteral Nutrition	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
sphingosine	Sphingosines	-0.028 (0.019)	0.146	0.015 (0.020)	0.444	-0.189 (0.096)	0.049
homocitrate	TCA Cycle	-0.054 (0.088)	0.539	0.057 (0.115)	0.621	-0.360 (0.167)	0.031
kynurenine	Tryptophan Metabolism	-0.093 (0.088)	0.293	0.064 (0.115)	0.578	-0.405 (0.161)	0.012
tryptophan		-0.197 (0.126)	0.118	0.079 (0.160)	0.622	-0.499 (0.230)	0.030

Beta beta-coefficient; *SE* standard error; *p* *p*-value

Analysis used logistic regression for the TOLSURF and PROP cohorts separately, stratified by nutrition source (parenteral vs enteral nutrition), adjusting for birth weight, sex, self-reported maternal race/ethnicity, and corticosteroids (>1 day), followed by a meta-analysis. A negative effect size/unit change (beta) indicates metabolites of lower abundance in BPD infants. Metabolites are grouped by sub-pathway. Sub-pathways and metabolites are listed alphabetically. Complete results are provided in the supplementary material.

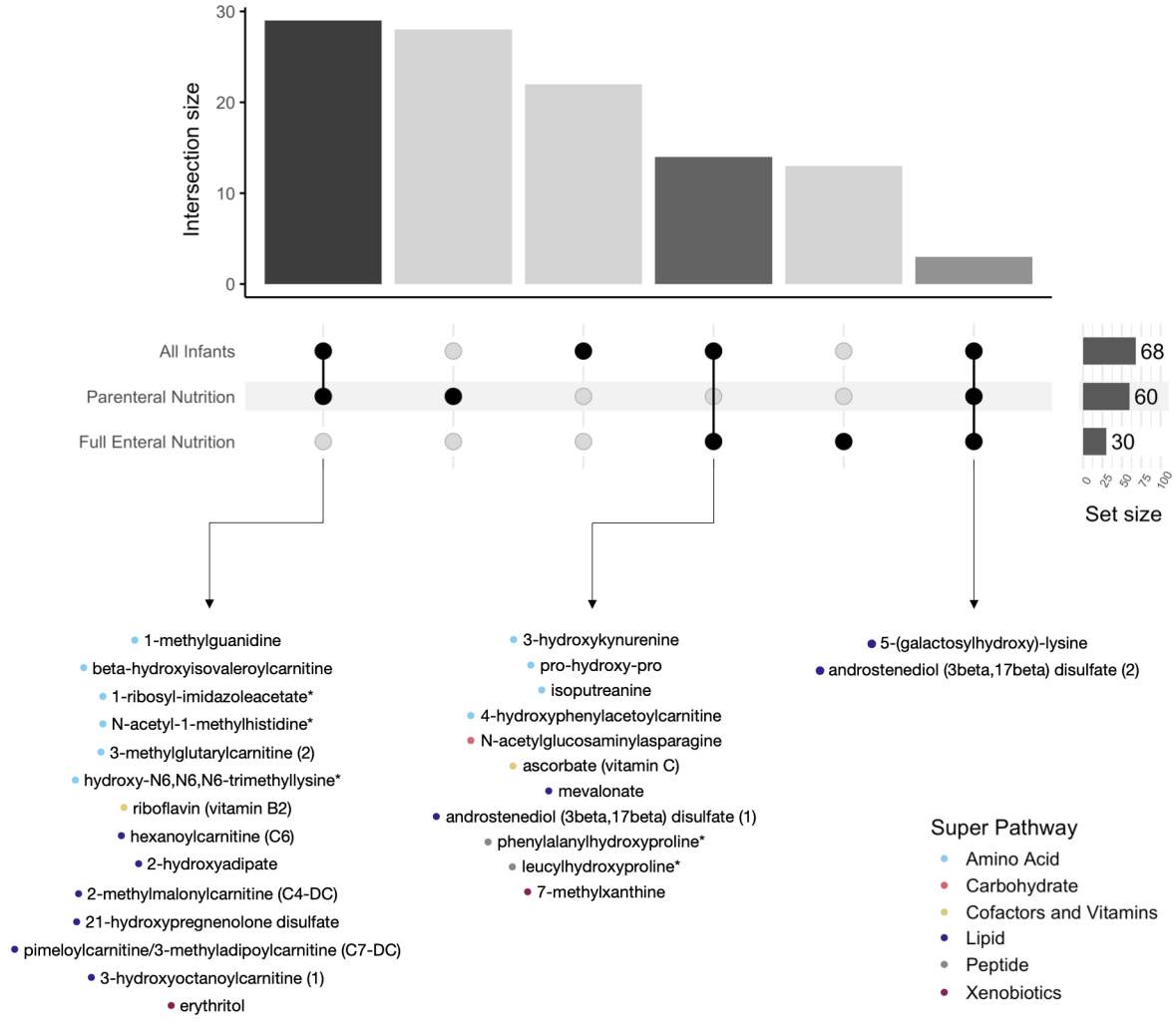

Figure 2.5. Differential metabolites at time point 2 across all infants and by nutrition source. Bars indicate the number of differential metabolites identified in each analysis (set size). Intersection bars show the number of metabolites shared across analyses. Differential metabolites of known identity in more than one analysis are listed below their corresponding intersections. Each metabolite is accompanied by a colored dot corresponding to its super pathway. The majority of differential metabolites within all infants at time point 2 are also differential within sub-analyses of infants by nutrition modality, including amino acids and lipids.

Pathway Enrichment Analysis

We performed pathway enrichment analyses for each MWAS using annotations provided by Metabolon. A total of 71 and 73 sub-pathways with ≥2 metabolites were evaluated for enrichment at time points 1 and 2, respectively, which belonged to 8 super pathways. At time point

1, differential metabolites were enriched for the Lactoyl Amino Acid sub-pathway (p=0.005, Table 2.6) within the Amino Acid super pathway. At time point 2, differential metabolites were significantly enriched in Fatty Acid Metabolism: Acyl Carnitine, Medium Chain (p=0.002), and Xanthine Metabolism (p=0.0006) (Table 2.6). For infants on parenteral nutrition at time point 2, differential metabolites were enriched in two lipid sub-pathways involved in fatty acid metabolism (Acyl Carnitine/Dicarboxylate [p=0.0002], and Acyl Carnitine/Hydroxy [p=0.0002]), and Vitamin B6 Metabolism (p=0.02) (Table 2.6). The Cofactors and Vitamins super-pathway was enriched for infants on parenteral nutrition at time point 2 (p=0.02). No other analyses showed enrichment of super-pathways. For infants on full enteral nutrition at time point 2, differential metabolites were enriched for Tryptophan Metabolism (p=0.03) and Androgenic Steroids (p=0.04) (Table 2.6). Differential metabolites identified in the fold-change comparison were enriched in Secondary Bile Acid Metabolism (p=0.008).

Table 2.6. Metabolomic sub-pathway enrichment analyses for biochemicals whose levels associate with BPD at $p<0.05$.

Time Point	Sub-Pathway	Enrichment Ratio	*p*	FDR
Time Point 1	Lactoyl Amino Acid	16.4	0.005	0.4
Time Point 2	Xanthine Metabolism	6.3	0.0006	0.04
	Fatty acid metabolism (Acyl Carnitine, Medium Chain)	9.1	0.002	0.09
Parenteral Nutrition, Time Point 2	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	16.6	0.0002	0.007
	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	16.6	0.0002	0.007
	Vitamin B6 Metabolism	8.3	0.02	0.5
Full Enteral Nutrition, Time Point 2	Tryptophan Metabolism	4.1	0.03	1
	Androgenic Steroids	5.8	0.04	1

p p-value; *FDR* false discovery rate within each group comparison.

Enrichment ratio was calculated as the proportion of metabolites at p<0.05 as compared to the total number of metabolites within the sub-pathway, divided by the proportion of total metabolites at p<0.05 over all metabolites evaluated across sub-pathways. An enrichment ratio >1 indicates that within the sub-pathway more metabolites than what would have been expected by chance have a p<0.05.

Discussion

In this study, we characterized the urinary metabolome of 314 extremely premature infants from two cohorts within the first four weeks of life, including infants with and without a later diagnosis of BPD at 36 weeks postmenstrual age. We detected over a thousand unique metabolites representing more than 50 metabolic pathways, highlighting the breadth of physiological processes captured through performing untargeted metabolomics in a relatively non-invasive biofluid. Cluster analyses across all metabolites revealed little to no separation between infants with and without BPD, suggesting a high degree of infant variability in the urinary metabolome in this large, population-based study. Levels of 89 unique individual metabolites were found to be associated with BPD across time points, for which the false discovery rate should be considered when weighing their individual relevance to BPD pathophysiology. However, taken together, “BPD differential” metabolites were significantly more likely to be at lower levels in the urine of infants with BPD as compared to infants without BPD, particularly during the fourth week of life and during the transition to enteral feeding. This was an unexpected finding and supports our hypothesis that infants who develop BPD are experiencing metabolic dysregulation early in life.

BPD differential metabolites were enriched in specific metabolic pathways, including lactoyl amino acids in the second week of life, and xanthine and fatty acid metabolism in the fourth week of life. At the earlier time point, we observed higher levels of two lactoyl amino acid metabolites in infants who developed BPD (N-lactoyl isoleucine and N-lactoyl leucine), leading to modest enrichment of the Lactoyl Amino Acid sub-pathway. N-lactoyl-amino acids are involved in a number of processes, including energy metabolism and sleep patterns,[26] and higher levels have been associated with a variety of other diseases, such as gestational diabetes[27] and retinopathy of prematurity.[28] Elevated levels in infants who develop BPD could indicate greater early postnatal

acidosis and a shift toward anaerobic metabolism, leading to increased lung injury and poor respiratory outcomes. Previous studies have also found an association between BPD and levels of urinary lactate, which is a precursor of lactoyl-amino acids, including the two we identified to be at higher levels in infants with BPD. In one study, lactate was found to be higher in infants with BPD,[15] whereas in another it was found to be lower in infants with BPD.[16] While we find no association between urinary lactate levels and BPD, we note that prior studies involving a smaller number of infants used a method that has lower sensitivity and is prone to degradation of lactoyl amino acids into lactate (hydrogen nuclear magnetic resonance, or H-NMR),[15,16] which may in part explain these discrepancies.

During the fourth week of life, levels of several caffeine metabolites (methylxanthines, methylurates) and acylcarnitines were lower in infants with BPD, with enrichment of the respective sub-pathways. Caffeine, an adenosine receptor antagonist commonly prescribed in the NICU to prevent apnea of prematurity, stimulates breathing, the central nervous system, and the cardiovascular system[29] and has been suggested to reduce the incidence of BPD.[30] Caffeine metabolism in preterm infants is limited by immature hepatic enzymes, while clearance is constrained by immature renal function, both of which are influenced by gestational age, post-conception age, and feeding modality.[29] Thus, lower levels of caffeine metabolites in infants who develop BPD may be attributed to these developmental limitations that reduce caffeine metabolism and renal excretion, as well as differences in dosing regimens based on body weight, caffeine blood levels, and apneic frequency. The use of caffeine was similar between infants with and without BPD, but data on dosing were not collected.

The enrichment in fatty acid metabolism was more pronounced in infants receiving parenteral nutrition as compared to infants receiving full enteral nutrition; however, longitudinal trajectories suggest the role of nutrition is complex and requires further study. Acylcarnitines and

carnitine are essential for energy production due to their crucial role in fatty acid oxidation, and thus they support metabolism and growth of the premature infant that could improve respiratory outcomes. Hydroxylated fatty acids, produced by tissues and some gut microbes, may also have anti-inflammatory effects and protect against lung damage caused by inflammation, which is also a hallmark of BPD. The involvement of carnitines in BPD has similarly been observed in prior studies that used different statistical approaches, metabolomics platforms, most commonly targeted metabolomics, and that included smaller cohorts varying by gestational age, postnatal age, modes of nutrition, and biofluid,[12,14,17–19] with conflicting directions of effect reported. Studies that collected samples during the first week of life or at 6 months corrected age found acylcarnitines to be higher in BPD infants.[12,14,19] Similarly, a study evaluating metabolic differences across nutritional stages reported that the direction of carnitine dysregulation changes during these transitional periods, with higher levels predominantly in infants who developed BPD.[18] In contrast, a cross-sectional study analyzing four time points found that most acylcarnitines, particularly short- and medium-chain carnitines, were lower in infants who developed BPD at postnatal day 28, but not at postnatal days 1, 7, or 42.[17] Our findings are consistent with those of the cross-sectional study in infants of similar postnatal age, including the replication of lower hexanoylcarnitine (C6) levels in infants who develop BPD and no differences at earlier time points. Collectively, our study, along with others, suggests that the relationship between fatty acid metabolism and BPD varies over time and by mode of nutrition (parenteral vs. enteral), a longitudinal pattern also observed for several carnitines, and that reduced fatty acid availability during the fourth week of life is either a marker or a contributor to BPD development.

We also observed lower levels of C and B-complex vitamins during the fourth week of life in infants who developed BPD, including vitamins B_1, B_6, and C among infants receiving parenteral nutrition with effects diminished among infants that transitioned to full enteral feeding. These findings are consistent with prior literature linking vitamin deficiencies to BPD.[31–33] Vitamin C is a

cofactor in reactions regulating hypoxia-induced transcription factors and enhancing immune and anti-inflammatory responses, and a deficiency has also been attributed to cases of pulmonary hypertension,[34] a common complication of BPD. B-complex vitamins support mitochondrial function[35] (vitamin B_2), the metabolism of carbohydrates[36] (vitamins B_1 and B_6), amino acids[36] (vitamin B_6), and lipids[36] (vitamin B_6); and gut health and development.[37–39] Thus, deficiencies in B vitamins may limit energy availability in the lung, which relies on the catabolism of glucose, proteins, and lipids.

Essential vitamins are obtained through nutrition and may have varying degrees of bioavailability and concentration in parenteral vs. enteral nutrition, prompting our sub-analyses comparing infants receiving parenteral nutrition (either exclusively, transitioning to enteral feeding, or full enteral feeds for <4 days) at time point 2 vs. those receiving full enteral nutrition for at least 4 days. It is important to note that during the feeding transition, infants routinely receive both parenteral and increasing amounts of enteral nutrition, and data on daily amounts were not available. Thus, lower circulating vitamin levels in infants who developed BPD could be due to impaired enteral absorption, immature gut, insufficient vitamin supplementation, and/or higher physiological demand in sicker preterm infants. Future interventions may include targeted nutritional supplementation for infants receiving parenteral nutrition, in particular vitamin C, B_1, B_2, and B_6 for infants with later or delayed transition to enteral feeds.

Vitamin deficiencies along with aberrant fatty acid metabolism could lead to suboptimal weight gain, reduced growth, and impaired lung repair, and warrant future investigations as to their role in BPD and their relationship with source of nutrition. The lung and gut of infants who develop BPD are less mature developmentally for their gestational age, resulting in an increased need for respiratory, nutritional, and metabolic support. Over the first few postnatal weeks, sicker infants may have increased work of breathing and resting energy expenditure, leading to slower

weight gain, less linear growth, and lower energy reserves. In addition, delays in maturational development and repair of lung structure and function could result in higher oxygen levels, increased ventilatory pressure requirements, and transient hypoxemia with associated oxidative stress and inflammation.

Strengths of our study include the multi-cohort design, including hundreds of infants from multiple centers at two time points, thereby capturing a greater degree of individual-level variation in the urinary metabolome of extremely premature infants. A potential weakness is the use of urine samples stored at -80°C for up to 10 years prior, which may have led to oxidation or degradation of certain important metabolites. In addition, our study lacked detailed information on the volume of enteral nutrition received, infant growth parameters, and caloric intake. Therefore, additional studies are needed to replicate our findings in a more contemporary cohort and to incorporate more detailed information on nutrition type.

In summary, we performed a metabolome-wide association study in extremely premature infants during the first few weeks of life and implicated specific metabolites and metabolic pathways in the development of BPD. Our results can inform future clinical trials to test interventions involving nutritional supplementation spanning lipids, essential amino acids, and vitamins B and C. Future studies are also needed to establish a causal relationship between metabolic dysregulation and the development of BPD, including the role of nutrition and nutrient absorption.

References

1. Cheong, J. L. Y. & Doyle, L. W. An update on pulmonary and neurodevelopmental outcomes of bronchopulmonary dysplasia. Seminars in Perinatology 42, 478–484 (2018).

2. Jensen, E. A. et al. Severity of Bronchopulmonary Dysplasia Among Very Preterm Infants in the United States. Pediatrics 148, e2020030007 (2021).

3. Ibrahim, J. & Bhandari, V. The definition of bronchopulmonary dysplasia: an evolving dilemma. Pediatr Res 84, 586–588 (2018).

4. Thekkeveedu, R. K., Guaman, M. C. & Shivanna, B. Bronchopulmonary dysplasia: A Review of Pathogenesis and Pathophysiology. Respir Med 132, 170–177 (2017).

5. Lewis, T. R. *et al.* Association of Racial Disparities With In-Hospital Outcomes in Severe Bronchopulmonary Dysplasia. *JAMA Pediatrics* **176**, 852–859 (2022).

6. Ambalavanan, N. *et al.* Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure. *J Perinatol* **28**, 420–426 (2008).

7. Hennelly, M., Greenberg, R. G. & Aleem, S. An Update on the Prevention and Management of Bronchopulmonary Dysplasia. *Pediatric Health, Medicine and Therapeutics* **12**, 405–419 (2021).

8. Principi, N., Di Pietro, G. M. & Esposito, S. Bronchopulmonary dysplasia: clinical aspects and preventive and therapeutic strategies. *J Transl Med* **16**, 36 (2018).

9. Qiu, S. *et al.* Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Sig Transduct Target Ther* **8**, 1–37 (2023).

10. You, Y. *et al.* Early metabolic markers as predictors of respiratory complications in preterm infants with bronchopulmonary dysplasia. *Early Human Development* **190**, 105950 (2024).

11. Course, C. W. *et al.* Evidence of abnormality in glutathione metabolism in the airways of preterm born children with a history of bronchopulmonary dysplasia. *Sci Rep* **13**, 19465 (2023).

12. Chiu, C.-Y. *et al.* Metabolomic Analysis Reveals the Association of Severe Bronchopulmonary Dysplasia with Gut Microbiota and Oxidative Response in Extremely Preterm Infants. *Metabolites* **14**, 219 (2024).

13. Sun, H. *et al.* Targeted blood metabolomics in infants with bronchopulmonary dysplasia. Preprint at https://doi.org/10.21203/rs.3.rs-4544343/v1 (2024).

14. La Frano, M. R. *et al.* Umbilical cord blood metabolomics reveal distinct signatures of dyslipidemia prior to bronchopulmonary dysplasia and pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L870–L881 (2018).

15. Fanos, V. *et al.* Urinary metabolomics of bronchopulmonary dysplasia (BPD): preliminary data at birth suggest it is a congenital disease. *The Journal of Maternal-Fetal & Neonatal Medicine* **27**, 39–45 (2014).

16. Pintus, M. C. *et al.* Urinary 1H-NMR Metabolomics in the First Week of Life Can Anticipate BPD Diagnosis. *Oxid Med Cell Longev* **2018**, 7620671 (2018).

17. Ye, C. *et al.* Progressive Metabolic Abnormalities Associated with the Development of Neonatal Bronchopulmonary Dysplasia. *Nutrients* **14**, 3547 (2022).

18. Wang, L., Zhong, W. H., Liu, D. Y., Shen, H. Q. & He, Z. J. Metabolic analysis of infants with bronchopulmonary dysplasia under early nutrition therapy: An observational cohort study. *Exp Biol Med (Maywood)* **247**, 470–479 (2022).

19. Piersigilli, F. *et al.* Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics* **15**, 20 (2019).

20. Malave-Mendez, T. *et al.* Metabolomic Fingerprint of Bronchopulmonary Dysplasia in Extremely Premature Infants. *Am J Respir Crit Care Med* **211**, A5432–A5432 (2025).

21. Ballard, R. A. *et al.* Inhaled Nitric Oxide in Preterm Infants Undergoing Mechanical Ventilation. *New England Journal of Medicine* **355**, 343–353 (2006).

22. Guardado, M. *et al.* The Urinary Metabolomic Fingerprint in Extremely Preterm Infants on Total Parenteral Nutrition vs. Enteral Feeds. *Metabolites* **13**, 971 (2023).

23. Torgerson, D. *et al.* The hydrocortisone-responsive urinary metabolome of premature infants. *Pediatr Res* **94**, 1317–1326 (2023).

24. Welcome to Python.org. *Python.org* https://www.python.org/ (2026).

25. R: The R Project for Statistical Computing. https://www.r-project.org/.

26. Naja, K., Hedaya, L., Elashi, A. A., Rizzo, M. & Elrayess, M. A. N-Lactoyl Amino Acids: Emerging Biomarkers in Metabolism and Disease. *Diabetes/Metabolism Research and Reviews* **41**, e70060 (2025).

27. Harville, E. W. *et al.* Branched-chain amino acids, history of gestational diabetes, and breastfeeding: The Bogalusa Heart Study. *Nutrition, Metabolism and Cardiovascular Diseases* **30**, 2077–2084 (2020).

28. Yang, Y. *et al.* Comparative Analysis Reveals Novel Changes in Plasma Metabolites and Metabolomic Networks of Infants With Retinopathy of Prematurity. *Invest. Ophthalmol. Vis. Sci.* **63**, 28 (2022).

29. Abdel-Hady, H., Nasef, N., Shabaan, A. E. & Nour, I. Caffeine therapy in preterm infants. *World J Clin Pediatr* **4**, 81–93 (2015).

30. Ma, J., Chen, L., Mu, K., Tang, S. & Shi, Y. Early caffeine therapy decreases bronchopulmonary dysplasia but might increase mortality in preterm infants? a systematic review and meta-analysis. *Front Pediatr* **13**, 1528054 (2025).

31. Park, H. W. *et al.* Association between vitamin D level and bronchopulmonary dysplasia: A systematic review and meta-analysis. *PLoS One* **15**, e0235332 (2020).

32. Ge, H. *et al.* The association of vitamin D and vitamin E levels at birth with bronchopulmonary dysplasia in preterm infants. *Pediatric Pulmonology* **56**, 2108–2113 (2021).

33. Verma, R. P., McCulloch, K. M., Worrell, L. & Vidyasagar, D. Vitamin A deficiency and severe bronchopulmonary dysplasia in very low birthweight infants. *Am J Perinatol* **13**, 389–393 (1996).

34. Hemilä, H. & de Man, A. M. E. Vitamin C deficiency can lead to pulmonary hypertension: a systematic review of case reports. *BMC Pulm Med* **24**, 140 (2024).

35. Balasubramaniam, S. & Yaplito-Lee, J. Riboflavin metabolism: role in mitochondrial function. *jtgg* **4**, 285–306 (2020).

36. Stach, K., Stach, W. & Augoff, K. Vitamin B6 in Health and Disease. *Nutrients* **13**, 3229 (2021).

37. Wibowo, S. & Pramadhani, A. Vitamin B and Vitamin E - Pleiotropic and Nutritional Benefits. in *Vitamin B, Role of Gut Microbiota and Gut Health* (IntechOpen, 2024).

38. Yates, C. A., Evans, G. S., Pearson, T. & Powers, H. J. Absence of luminal riboflavin disturbs early postnatal development of the gastrointestinal tract. *Dig Dis Sci* **48**, 1159–1164 (2003).

39. Park, J. *et al.* Dietary Vitamin B1 Intake Influences Gut Microbial Community and the Consequent Production of Short-Chain Fatty Acids. *Nutrients* **14**, 2078 (2022).

40. Ballard, R. A. *et al.* Randomized Trial of Late Surfactant Treatment in Ventilated Preterm Infants Receiving inhaled Nitric Oxide. *J Pediatr* **168**, 23-29.e4 (2016).

41. Poindexter, B. B. *et al.* Comparisons and Limitations of Current Definitions of Bronchopulmonary Dysplasia for the Prematurity and Respiratory Outcomes Program. *Annals ATS* **12**, 1822–1830 (2015).

42. Ballard, P. L. *et al.* Blood metabolomics in infants enrolled in a dose escalation pilot trial of budesonide in surfactant. *Pediatr Res* **90**, 784–794 (2021).

43. Storey, J., Bass, A., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. https://doi.org/10.18129/B9.bioc.qvalue (2025).

Supplementary Material

Extended Methods

Study Description

Infants included in our study were from the Trial of Late Surfactant for Prevention of Bronchopulmonary Dysplasia (TOLSURF, ClinicalTrials.gov, NCT01022580) and the Prematurity and Respiratory Outcomes Program (PROP, NCT01435187). Clinical data were obtained for both cohorts, including gestational age, birthweight, maternal race/ethnicity, use of corticosteroids, mode of feeding (parenteral vs. enteral), and respiratory parameters, including Respiratory Severity Scores (RSS = FiO2 × Mean Airway Pressure). The research protocol for both studies was approved by the institutional review boards of the participating institutions, and a parent of each infant provided written informed consent.

A total of 171 infants from TOLSURF were included in the current study, and 143 infants from PROP. Infants were selected from available samples and attempts were made to balance by bronchopulmonary dysplasia (BPD) at 36 weeks gestational age, and by maternal Black, Hispanic, and non-Hispanic White race/ethnicity. TOLSURF was a double-blinded, randomized, sham-controlled trial to evaluate the effect of late surfactant on BPD and included 511 infants across 25 US centers enrolled between 2010 and 2013.[1] Enrollment criteria included intubation and mechanical ventilation between 7 and 14 days of age and gestational age between 23-28 weeks. Infants received iNO according to the protocol followed in the Nitric Oxide (to Prevent) Chronic Lung Disease (NO CLD) trial, and BPD was assessed at 36 weeks post-menstrual age by physiological testing.[2] There was no significant difference in BPD incidence between placebo and surfactant-treated groups at 36 weeks. PROP was an observational study of 835 infants <29 weeks of gestational age from 8 US centers enrolled between 2010 and 2013.[3] BPD was

assessed at 36 weeks post-menstrual age by physiological testing using the same definition as for TOLSURF.

Sample Collection

Urine samples were collected from infants by placing cotton balls into the diaper for 4-8 hours and then transferred to a tube and frozen at −70 °C. Samples were selected from two time points for each infant: between postnatal days 5-15 (time point 1; mean/sd 8±2) and 23-32 (time point 2; mean/sd 28±2), for a total of 628 urine samples. The total number of infants in our study was 314, including 178 with BPD and 136 without BPD (TOLSURF: N=106 with BPD and 65 without BPD; PROP: N=72 with BPD and 71 without BPD). Infants were selected based on survival at 40 weeks gestational age and availability of urine samples. In TOLSURF, infants were categorized as being on parenteral nutrition at time point 2 if they were receiving or had received parenteral nutrition within four days prior to urine sample collection, or as being on full enteral nutrition if they had not received parenteral nutrition for at least 4 days prior. In PROP, infants were categorized as being on parenteral nutrition if they had received either no or partial enteral feeds within 4 days prior to sample collection, and on enteral nutrition if they had received full enteral feeds within the same timeframe.

Metabolomics Profiling

Untargeted metabolomic assays were performed using Ultra High-Performance Liquid Chromatography-Tandem Mass Spectrometry (UHPLC/MS-MS) by Metabolon Inc. as previously described.[4–6] Metabolic profiling was conducted separately for each cohort, with anchor samples included to facilitate the comparison of metabolite abundance between cohorts. Each analytic batch included a set of internal standards to adjust values for instrument variability as needed. Urine samples were normalized to osmolality, and metabolites with >30% missing values (below the limit of detection) within TOLSURF or PROP were removed from further analysis. Levels of

missing individual metabolites were imputed to the minimum detectable level, and data were normalized to the median value.

Statistical Analyses

Principal component analysis (PCA) was performed to evaluate global differences between the urinary metabolome of infants with and without BPD at each time point and by nutrition source. A Mann-Whitney U-test was performed within each analysis to assess differences in the distributions of the first two principal components by BPD status. These analyses were performed using R version 4.5.1 (R Foundation for Statistical Computing).

A metabolome-wide association study (MWAS) was performed using multivariate logistic regression to test the association between BPD and levels of individual metabolites using the *statsmodels* package[7] in Python 3.11.4. Covariates in the model included birth weight, sex, self-identified maternal race/ethnicity, and whether an infant was receiving corticosteroids for more than 1 day prior to urine collection. Nutrition source was included as a covariate only for time point 2. Analyses were also performed on the fold-change, calculated using the ratio of metabolite levels at time point 2 to those at time point 1. To account for study-specific and experimental batch effects, analyses were done within each cohort separately and combined using a fixed-effects meta-analysis with the *metafor* package in R 4.5.1.[8] Sub-analyses were performed, separating infants by nutrition source at time point 2 to evaluate its role as a potential confounder and effect modifier. Metabolites were defined as “differential metabolites” when associated with BPD at a threshold of $p<0.05$, given the high degree of correlation within our dataset.

Longitudinal data for differential metabolites from a prior study[5] were aligned to days relative to discontinuation of parenteral nutrition and visualized using time course plots with locally weighted scatterplot smoothing by BPD status. Data included untargeted metabolomics of 302 urine samples from 28 infants in TOLSURF with ≥5 time points (mean/sd 11±4) over ≥19 days

(mean/sd 35±8). In this cohort, 17 (60.7%) of the infants developed BPD, and 11 (39.3%) did not develop BPD; 25 of the 28 infants were also included in the two time point study.

Pathway enrichment analysis was performed on differential metabolites that were associated with BPD at $p<0.05$ for each of the MWAS comparisons performed. Metabolites of known identity from Metabolon (as of April 2025) were annotated to 8 super-pathways, including 88 sub-pathways at time point 1, and 90 sub-pathways at time point 2. Sub-pathways with less than 2 annotated metabolites were removed from the analysis. MSEA was performed for each super-pathway and sub-pathway using a Fisher's exact test, adjusting for multiple comparisons by inferring the FDR using qvalue. The enrichment ratio was calculated as the proportion of metabolites at $p<0.05$, compared to the total number of metabolites within the sub-pathway, divided by the proportion of total metabolites at $p<0.05$ over all metabolites evaluated across sub-pathways and super-pathways. Multiple comparisons were accounted for by inferring the false discovery rate (FDR) using the *qvalue* package from Bioconductor in R 4.5.1.[9]

References

1. Ballard, R. A. *et al.* Randomized Trial of Late Surfactant Treatment in Ventilated Preterm Infants Receiving inhaled Nitric Oxide. *J. Pediatr.* **168**, 23-29.e4 (2016).

2. Ballard, R. A. *et al.* Inhaled Nitric Oxide in Preterm Infants Undergoing Mechanical Ventilation. *N. Engl. J. Med.* **355**, 343–353 (2006).

3. Poindexter, B. B. *et al.* Comparisons and Limitations of Current Definitions of Bronchopulmonary Dysplasia for the Prematurity and Respiratory Outcomes Program. *Ann. Am. Thorac. Soc.* **12**, 1822–1830 (2015).

4. Guardado, M. *et al.* The Urinary Metabolomic Fingerprint in Extremely Preterm Infants on Total Parenteral Nutrition vs. Enteral Feeds. *Metabolites* **13**, 971 (2023).

5. Torgerson, D. *et al.* The hydrocortisone-responsive urinary metabolome of premature infants. *Pediatr. Res.* **94**, 1317–1326 (2023).

6. Ballard, P. L. *et al.* Blood metabolomics in infants enrolled in a dose escalation pilot trial of budesonide in surfactant. *Pediatr. Res.* **90**, 784–794 (2021).

7. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *SciPy 2010* https://doi.org/10.25080/Majora-92bf1922-011 (2010) doi:10.25080/Majora-92bf1922-011.

8. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).

9. Storey, J., Bass, A., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. https://doi.org/10.18129/B9.bioc.qvalue (2025).

Supplementary Figure

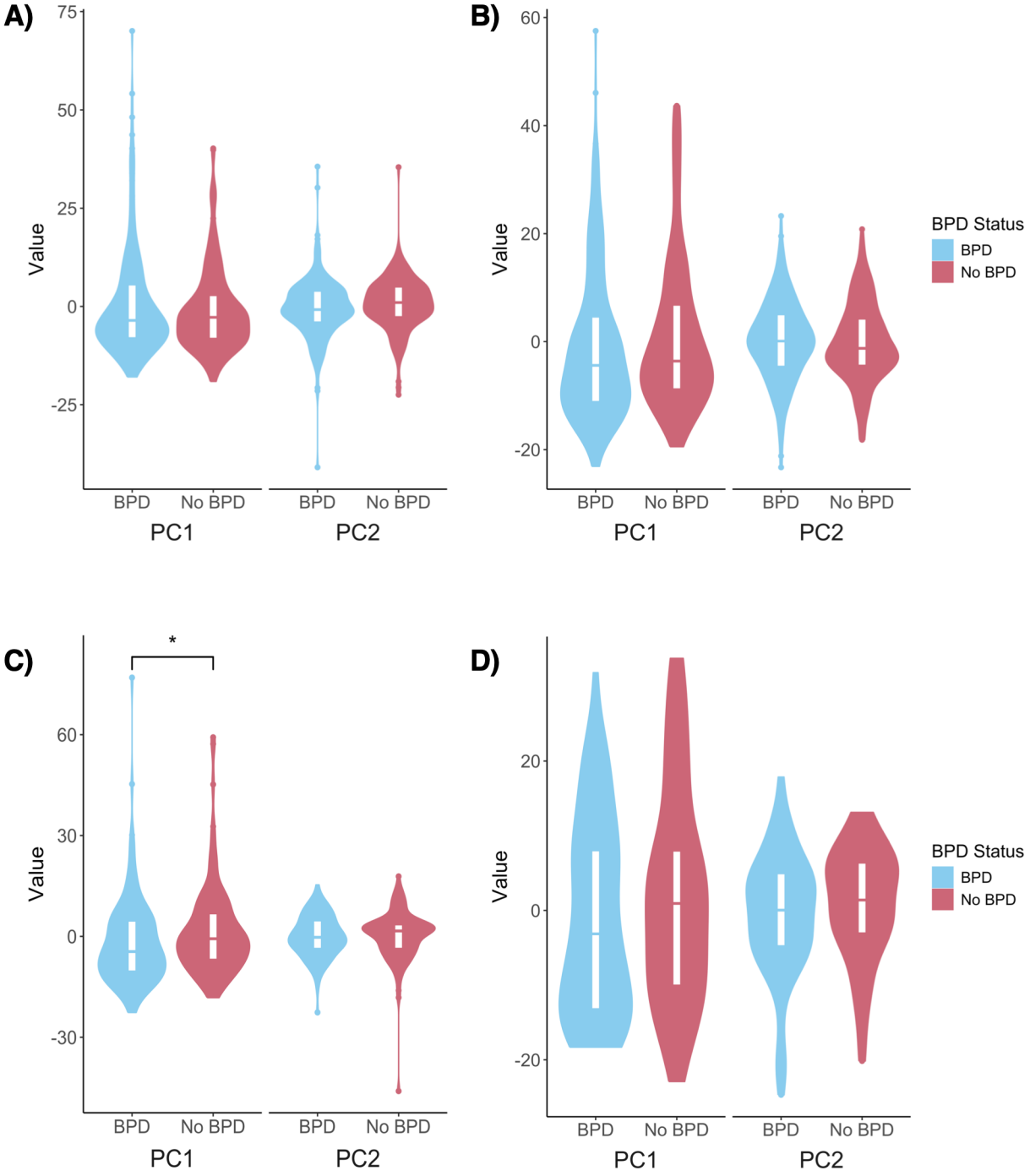

Figure S2.1. Violin plots of principal component 1 (PC1) and principal component 2 (PC2) distributions by BPD status.

Comparisons are shown for: (A) all infants at time point 1 (postnatal days 5-15), (B) all infants at time point 2 (postnatal days 23-32), (C) infants on parenteral nutrition at time point 2, and (D) infants on enteral nutrition at time point 2.

Table S2.1. Summary statistics of differential metabolites (p<0.05) at time point 1 (postnatal days 5-15)

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
sucrose	Disaccharides and Oligosaccharides	0.074 (0.156)	0.638	-0.081 (0.030)	0.007	-0.076 (0.029)	0.010
4-methylnonanoylcarnitine	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	0.021 (0.104)	0.838	-0.244 (0.087)	0.005	-0.136 (0.067)	0.042
methyl vanillate sulfate	Food Component/Plant	0.137 (0.074)	0.064	0.067 (0.048)	0.155	0.088 (0.040)	0.028
menthol glucuronide	Food Component/Plant	-0.079 (0.191)	0.680	-0.175 (0.081)	0.031	-0.160 (0.075)	0.032
5-aminolevulinate	Hemoglobin and Porphyrin Metabolism	0.526 (0.326)	0.106	0.756 (0.422)	0.073	0.612 (0.258)	0.018
N-lactoyl isoleucine	Lactoyl Amino Acid	0.389 (0.265)	0.143	0.546 (0.398)	0.171	0.437 (0.221)	0.048
N-lactoyl leucine	Lactoyl Amino Acid	0.469 (0.290)	0.106	0.458 (0.398)	0.250	0.465 (0.234)	0.047
beta-hydroxyisovaleroylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.093 (0.248)	0.707	-0.512 (0.214)	0.017	-0.332 (0.162)	0.040
pipecolate	Lysine Metabolism	0.405 (0.247)	0.100	0.599 (0.377)	0.112	0.463 (0.206)	0.025
glucuronide of C10H18O2 (11)*	Partially Characterized Molecules	0.443 (0.216)	0.040	0.276 (0.206)	0.181	0.356 (0.149)	0.017
glucuronide of C10H18O2 (12)*	Partially Characterized Molecules	0.284 (0.154)	0.064	0.182 (0.213)	0.394	0.249 (0.125)	0.046
guanine	Purine Metabolism, Guanine containing	0.144 (0.161)	0.371	0.904 (0.295)	0.002	0.318 (0.141)	0.024
3-hydroxykynurenine	Tryptophan Metabolism	0.349 (0.175)	0.045	0.388 (0.312)	0.214	0.359 (0.152)	0.019
tyrosine	Tyrosine Metabolism	0.163 (0.133)	0.222	1.253 (0.421)	0.003	0.262 (0.127)	0.039
caffeine	Xanthine Metabolism	-0.385 (0.205)	0.060	-0.226 (0.237)	0.342	-0.317 (0.155)	0.041

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
X-11381		-0.072 (0.059)	0.218	-0.115 (0.056)	0.040	-0.095 (0.041)	0.019
X-13431		-0.055 (0.143)	0.702	-0.540 (0.168)	0.001	-0.258 (0.109)	0.018
X-17707		0.075 (0.056)	0.185	0.081 (0.049)	0.100	0.078 (0.037)	0.035
X-17763		0.060 (0.057)	0.295	0.080 (0.047)	0.091	0.072 (0.036)	0.049
X-24328		0.557 (0.254)	0.028	0.122 (0.335)	0.715	0.398 (0.202)	0.049
X-24359		0.489 (0.196)	0.012	0.252 (0.211)	0.231	0.380 (0.143)	0.008
X-24432		0.109 (0.162)	0.502	0.400 (0.147)	0.007	0.268 (0.109)	0.014
X-24801		-0.091 (0.184)	0.621	-0.436 (0.182)	0.016	-0.266 (0.129)	0.040

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Table S2.2. Summary statistics of differential metabolites (p<0.05) at time point 2 (postnatal days 23-32)

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	p	Beta (SE)	p	Beta (SE)	p
N-acetylglucosaminylasparagine	Aminosugar Metabolism	-0.470 (0.271)	0.084	-0.385 (0.335)	0.251	-0.436 (0.211)	0.039
androstenediol (3beta,17beta) disulfate (1)	Androgenic Steroids	-0.161 (0.134)	0.227	-0.231 (0.141)	0.102	-0.194 (0.097)	0.045
androstenediol (3beta,17beta) disulfate (2)	Androgenic Steroids	-0.557 (0.239)	0.02	-0.391 (0.231)	0.091	-0.471 (0.166)	0.005
ascorbate (vitamin C)	Ascorbate and Aldarate Metabolism	-0.025 (0.010)	0.013	-0.032 (0.056)	0.571	-0.025 (0.010)	0.011
diaminopimelate	Bacterial/Fungal	-0.194 (0.130)	0.134	-0.159 (0.120)	0.186	-0.175 (0.088)	0.047
sulfate*	Chemical	-0.533 (0.301)	0.077	-0.354 (0.321)	0.271	-0.449 (0.220)	0.041
3-hydroxypyridine sulfate	Chemical	-0.113 (0.071)	0.113	-0.168 (0.115)	0.143	-0.128 (0.061)	0.034
phenylalanylhydroxyproline*	Dipeptide	-0.286 (0.111)	0.01	-0.079 (0.111)	0.479	-0.183 (0.078)	0.02
leucylhydroxyproline*	Dipeptide Derivative	-0.205 (0.092)	0.026	-0.089 (0.116)	0.443	-0.160 (0.072)	0.026
pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	-0.206 (0.120)	0.086	-0.197 (0.117)	0.093	-0.201 (0.084)	0.016
3-hydroxyoctanoylcarnitine (1)	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.053 (0.101)	0.604	-0.440 (0.143)	0.002	-0.182 (0.083)	0.027
hexanoylcarnitine (C6)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.079 (0.058)	0.173	-0.236 (0.094)	0.012	-0.122 (0.049)	0.013
octanoylcarnitine (C8)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.054 (0.037)	0.143	-0.861 (0.226)	0.0001	-0.075 (0.037)	0.039
2-methylhexanoylcarnitine	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.056 (0.061)	0.362	-0.428 (0.147)	0.004	-0.111 (0.056)	0.049
2-methylmalonylcarnitine (C4-DC)	Fatty Acid Metabolism (also BCAA Metabolism)	-0.280 (0.242)	0.247	-0.902 (0.369)	0.014	-0.467 (0.202)	0.021

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
2-hydroxyadipate	Fatty Acid, Dicarboxylate	-0.447 (0.277)	0.107	-0.653 (0.359)	0.069	-0.524 (0.219)	0.017
erythritol	Food Component/Plant	-0.595 (0.301)	0.048	-0.442 (0.295)	0.134	-0.517 (0.211)	0.014
1-methylguanidine	Guanidino and Acetamido Metabolism	-0.128 (0.120)	0.285	-0.439 (0.174)	0.012	-0.228 (0.099)	0.021
4-imidazoleacetate	Histidine Metabolism	0.426 (0.265)	0.108	0.289 (0.224)	0.197	0.346 (0.171)	0.043
1-ribosyl-imidazoleacetate*	Histidine Metabolism	0.374 (0.192)	0.052	0.251 (0.162)	0.122	0.302 (0.124)	0.015
N-acetyl-1-methylhistidine*	Histidine Metabolism	-0.351 (0.182)	0.053	-0.369 (0.271)	0.174	-0.357 (0.151)	0.018
beta-hydroxyisovaleroylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.267 (0.225)	0.236	-1.053 (0.322)	0.001	-0.525 (0.185)	0.004
tiglylcarnitine (C5:1-DC)	Leucine, Isoleucine and Valine Metabolism	-0.135 (0.184)	0.463	-0.980 (0.310)	0.002	-0.356 (0.158)	0.025
3-methylglutarylcarnitine (2)	Leucine, Isoleucine and Valine Metabolism	-0.078 (0.123)	0.523	-0.389 (0.140)	0.006	-0.213 (0.092)	0.021
methylsuccinoylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.095 (0.082)	0.246	-0.147 (0.073)	0.045	-0.124 (0.055)	0.024
pipecolate	Lysine Metabolism	0.395 (0.193)	0.041	0.198 (0.188)	0.293	0.294 (0.135)	0.029
5-(galactosylhydroxy)-lysine	Lysine Metabolism	-0.619 (0.243)	0.011	-0.442 (0.286)	0.123	-0.545 (0.185)	0.003
hydroxy-N6,N6,N6-trimethyllysine*	Lysine Metabolism	-0.593 (0.285)	0.038	-0.555 (0.356)	0.119	-0.578 (0.223)	0.009
taurine	Methionine, Cysteine, SAM and Taurine Metabolism	-0.097 (0.114)	0.396	-0.209 (0.109)	0.054	-0.155 (0.079)	0.048
mevalonate	Mevalonate Metabolism	-0.385 (0.157)	0.014	-0.328 (0.165)	0.046	-0.358 (0.114)	0.002
N,N-dimethyl-pro-pro	Modified Peptides	-0.648 (0.374)	0.083	-0.511 (0.406)	0.208	-0.585 (0.275)	0.033

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
isoputreanine	Polyamine Metabolism	-0.308 (0.239)	0.197	-0.639 (0.362)	0.078	-0.408 (0.199)	0.041
21-hydroxypregnenolone disulfate	Pregnenolone Steroids	-0.326 (0.180)	0.07	-0.263 (0.164)	0.109	-0.292 (0.121)	0.016
uridine	Pyrimidine Metabolism, Uracil containing	0.101 (0.106)	0.34	0.182 (0.094)	0.053	0.146 (0.070)	0.037
riboflavin (vitamin B2)	Riboflavin Metabolism	-0.494 (0.194)	0.011	-0.227 (0.183)	0.216	-0.353 (0.133)	0.008
5-hydroxyindoleacetate	Tryptophan Metabolism	-0.544 (0.233)	0.02	-0.129 (0.243)	0.595	-0.345 (0.168)	0.04
3-hydroxykynurenine	Tryptophan Metabolism	-0.447 (0.198)	0.024	-0.116 (0.091)	0.202	-0.174 (0.083)	0.035
4-hydroxyphenylacetoylcarnitine	Tyrosine Metabolism	-0.024 (0.014)	0.081	-0.032 (0.020)	0.112	-0.026 (0.011)	0.019
pro-hydroxy-pro	Urea cycle; Arginine and Proline Metabolism	-0.660 (0.269)	0.014	-0.190 (0.194)	0.328	-0.351 (0.157)	0.026
3-methylxanthine	Xanthine Metabolism	-0.050 (0.043)	0.24	-0.063 (0.035)	0.071	-0.058 (0.027)	0.032
7-methylxanthine	Xanthine Metabolism	-0.061 (0.057)	0.281	-0.145 (0.077)	0.061	-0.090 (0.046)	0.048
1,7-dimethylurate	Xanthine Metabolism	-0.158 (0.090)	0.079	-0.124 (0.076)	0.102	-0.138 (0.058)	0.017
1-methylurate	Xanthine Metabolism	-0.496 (0.241)	0.04	-0.173 (0.186)	0.354	-0.294 (0.147)	0.046
7-methylurate	Xanthine Metabolism	-0.058 (0.052)	0.267	-0.129 (0.063)	0.039	-0.087 (0.040)	0.03
X-12117		-0.258 (0.218)	0.236	-0.627 (0.279)	0.025	-0.398 (0.172)	0.021
X-12708		-0.340 (0.310)	0.272	-0.612 (0.312)	0.05	-0.475 (0.220)	0.031
X-12721		-0.523 (0.241)	0.03	-0.224 (0.270)	0.406	-0.391 (0.180)	0.03

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
X-12822		-0.717 (0.288)	0.013	-0.448 (0.378)	0.237	-0.618 (0.229)	0.007
X-12844		-0.118 (0.073)	0.106	-0.164 (0.066)	0.012	-0.143 (0.049)	0.003
X-13431		-0.055 (0.139)	0.691	-0.687 (0.222)	0.002	-0.233 (0.118)	0.048
X-15136		-0.488 (0.202)	0.016	-0.115 (0.248)	0.642	-0.339 (0.157)	0.03
X-17357		-0.083 (0.081)	0.302	-0.280 (0.097)	0.004	-0.164 (0.062)	0.008
X-18838		-0.268 (0.111)	0.016	-0.223 (0.141)	0.113	-0.251 (0.087)	0.004
X-21364		-0.328 (0.135)	0.015	-0.156 (0.149)	0.296	-0.251 (0.100)	0.012
X-22152		-0.260 (0.217)	0.23	-0.370 (0.198)	0.062	-0.320 (0.146)	0.029
X-23518		-0.276 (0.131)	0.035	-0.375 (0.233)	0.108	-0.300 (0.114)	0.009
X-23654		-0.330 (0.188)	0.079	-0.578 (0.239)	0.016	-0.424 (0.148)	0.004
X-23787		-0.332 (0.162)	0.04	-0.150 (0.083)	0.071	-0.188 (0.074)	0.011
X-23918		-0.188 (0.102)	0.065	-0.086 (0.076)	0.262	-0.123 (0.061)	0.045
X-23922		-0.218 (0.196)	0.267	-0.392 (0.148)	0.008	-0.328 (0.118)	0.006
X-24340		-0.280 (0.144)	0.052	-0.123 (0.111)	0.265	-0.182 (0.088)	0.039
X-24342		-0.211 (0.136)	0.12	-0.516 (0.194)	0.008	-0.311 (0.111)	0.005
X-24468		-0.288 (0.136)	0.034	-0.322 (0.123)	0.009	-0.307 (0.091)	0.001

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
X-24588		-0.097 (0.070)	0.162	-0.184 (0.088)	0.037	-0.131 (0.055)	0.017
X-24801		-0.227 (0.185)	0.22	-0.455 (0.193)	0.018	-0.336 (0.133)	0.012
X-25420		-0.083 (0.036)	0.022	-0.069 (0.072)	0.331	-0.080 (0.032)	0.013
X-25422		-0.303 (0.198)	0.125	-1.080 (0.314)	0.001	-0.524 (0.167)	0.002
X-25457		-0.383 (0.138)	0.006	-0.334 (0.159)	0.035	-0.362 (0.104)	0.001

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Table S2.3. Summary statistics of metabolites with fold-change associated with BPD (p<0.05)

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
threonate	Ascorbate and Aldarate Metabolism	-0.002 (0.002)	0.198	-0.001 (0.001)	0.106	-0.002 (0.001)	0.043
hexanoylcarnitine (C6)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.339 (0.168)	0.043	-0.150 (0.244)	0.540	-0.278 (0.138)	0.044
2-hydroxystearate	Fatty Acid, Monohydroxy	-0.043 (0.039)	0.273	-0.092 (0.048)	0.055	-0.063 (0.030)	0.039
2-hydroxybehenate	Fatty Acid, Monohydroxy	-0.009 (0.010)	0.362	-0.025 (0.010)	0.018	-0.017 (0.007)	0.021
(S)-a-amino-omega-caprolactam	Food Component/Plant	-0.264 (0.126)	0.037	-0.362 (0.541)	0.504	-0.269 (0.123)	0.029
carnosine	Histidine Metabolism	-0.028 (0.026)	0.277	-0.069 (0.030)	0.021	-0.046 (0.020)	0.02
hydantoin-5-propionate	Histidine Metabolism	-0.717 (0.710)	0.313	-0.209 (0.099)	0.034	-0.219 (0.098)	0.025
2,3-dihydroxy-3-methylvalerate	Leucine, Isoleucine and Valine Metabolism	-0.012 (0.008)	0.125	-0.028 (0.018)	0.121	-0.015 (0.007)	0.043
glutarylcarnitine (C5-DC)	Lysine Metabolism	-0.022 (0.017)	0.189	-0.012 (0.008)	0.109	-0.014 (0.007)	0.045
6-oxopiperidine-2-carboxylate	Lysine Metabolism	-0.032 (0.020)	0.113	-0.025 (0.020)	0.208	-0.029 (0.014)	0.044
N-acetyl-isoputreanine	Polyamine Metabolism	-0.104 (0.051)	0.040	-0.043 (0.038)	0.264	-0.065 (0.031)	0.033
5alpha-pregnan-3beta,20alpha-diol disulfate	Progestin Steroids	-0.030 (0.023)	0.187	-0.033 (0.021)	0.106	-0.032 (0.015)	0.037
taurocholenate sulfate*	Secondary Bile Acid Metabolism	-0.055 (0.034)	0.104	-0.067 (0.040)	0.095	-0.060 (0.026)	0.02
glycohyocholate	Secondary Bile Acid Metabolism	-0.217 (0.088)	0.014	-0.050 (0.059)	0.398	-0.102 (0.049)	0.039
tricarballylate	TCA Cycle	-0.133 (0.088)	0.131	-0.169 (0.130)	0.195	-0.145 (0.073)	0.048

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
gamma-CEHC sulfate	Tocopherol Metabolism	-0.227 (0.140)	0.105	-0.166 (0.122)	0.171	-0.192 (0.092)	0.036
3-indoxyl sulfate	Tryptophan Metabolism	-0.036 (0.022)	0.103	-0.026 (0.020)	0.193	-0.031 (0.015)	0.039
tryptophan betaine	Tryptophan Metabolism	-0.089 (0.064)	0.165	-0.090 (0.043)	0.036	-0.090 (0.036)	0.012
N-methylproline	Urea cycle; Arginine and Proline Metabolism	0.073 (0.057)	0.196	0.036 (0.020)	0.074	0.040 (0.019)	0.034
X-11612		-0.071 (0.037)	0.056	-0.099 (0.071)	0.16	-0.077 (0.033)	0.019
X-12117		-0.236 (0.139)	0.089	-0.344 (0.294)	0.242	-0.256 (0.126)	0.042
X-12844		-0.013 (0.021)	0.535	-0.017 (0.008)	0.044	-0.016 (0.008)	0.035
X-15486		-0.050 (0.029)	0.091	-0.024 (0.017)	0.152	-0.030 (0.014)	0.038
X-17357		-0.034 (0.040)	0.391	-0.055 (0.029)	0.059	-0.048 (0.024)	0.043
X-17674		-0.055 (0.036)	0.127	-0.215 (0.105)	0.041	-0.072 (0.034)	0.035
X-21815		-0.023 (0.011)	0.033	-0.009 (0.009)	0.345	-0.015 (0.007)	0.036
X-24669		-0.015 (0.009)	0.084	-0.012 (0.010)	0.239	-0.014 (0.007)	0.038
X-25422		-0.154 (0.075)	0.039	-0.804 (0.319)	0.012	-0.188 (0.073)	0.01

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Table S2.4. Summary statistics of differential metabolites (p<0.05) at time point 2 (postnatal days 23-32) in infants on parenteral nutrition

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
androstenediol (3beta,17beta) disulfate (2)	Androgenic Steroids	-0.481 (0.319)	0.132	-0.513 (0.318)	0.107	-0.497 (0.225)	0.027
dehydroascorbate	Ascorbate and Aldarate Metabolism	-0.168 (0.091)	0.064	-0.210 (0.141)	0.137	-0.181 (0.076)	0.018
ascorbic acid 3-sulfate*	Ascorbate and Aldarate Metabolism	-0.426 (0.211)	0.044	-0.369 (0.208)	0.076	-0.397 (0.148)	0.007
lactose	Disaccharides and Oligosaccharides	-0.257 (0.135)	0.057	-0.262 (0.162)	0.107	-0.259 (0.104)	0.012
adipoylcarnitine (C6-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	-0.235 (0.130)	0.070	-0.154 (0.121)	0.204	-0.192 (0.089)	0.030
suberoylcarnitine (C8-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	-0.154 (0.083)	0.064	-0.068 (0.064)	0.287	-0.100 (0.051)	0.048
pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)	Fatty Acid Metabolism (Acyl Carnitine, Dicarboxylate)	-0.258 (0.181)	0.153	-0.451 (0.188)	0.016	-0.351 (0.130)	0.007
(S)-3-hydroxybutyrylcarnitine	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.149 (0.114)	0.191	-0.773 (0.234)	0.001	-0.268 (0.102)	0.009
3-hydroxyhexanoylcarnitine (1)	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.072 (0.161)	0.657	-0.502 (0.189)	0.008	-0.253 (0.123)	0.039
3-hydroxyoctanoylcarnitine (1)	Fatty Acid Metabolism (Acyl Carnitine, Hydroxy)	-0.161 (0.205)	0.434	-0.950 (0.277)	0.001	-0.440 (0.165)	0.008
hexanoylcarnitine (C6)	Fatty Acid Metabolism (Acyl Carnitine, Medium Chain)	-0.109 (0.081)	0.178	-0.533 (0.191)	0.005	-0.174 (0.075)	0.020
hexanoylglycine	Fatty Acid Metabolism (Acyl Glycine)	-0.248 (0.122)	0.041	-0.122 (0.190)	0.521	-0.211 (0.102)	0.039
2-methylmalonylcarnitine (C4-DC)	Fatty Acid Metabolism (also BCAA Metabolism)	-0.642 (0.348)	0.065	-1.052 (0.451)	0.020	-0.795 (0.275)	0.004
malonylcarnitine	Fatty Acid Synthesis	-0.374 (0.296)	0.206	-0.732 (0.313)	0.019	-0.543 (0.215)	0.011

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
2-hydroxyadipate	Fatty Acid, Dicarboxylate	-0.736 (0.398)	0.064	-0.787 (0.451)	0.081	-0.758 (0.298)	0.011
dimethylmalonic acid	Fatty Acid, Dicarboxylate	-0.319 (0.457)	0.486	-0.876 (0.428)	0.041	-0.616 (0.313)	0.049
erythritol	Food Component/Plant	-0.837 (0.495)	0.090	-0.708 (0.398)	0.075	-0.759 (0.310)	0.014
piperidine	Food Component/Plant	-0.030 (0.021)	0.145	-0.057 (0.036)	0.112	-0.037 (0.018)	0.040
furaneol sulfate	Food Component/Plant	-0.363 (0.179)	0.042	-0.129 (0.124)	0.297	-0.205 (0.102)	0.044
1-methylguanidine	Guanidino and Acetamido Metabolism	-0.146 (0.175)	0.402	-1.039 (0.328)	0.002	-0.343 (0.154)	0.026
1-ribosyl-imidazoleacetate*	Histidine Metabolism	0.406 (0.239)	0.089	0.484 (0.263)	0.066	0.441 (0.177)	0.013
N-acetyl-1-methylhistidine*	Histidine Metabolism	-0.369 (0.231)	0.109	-0.441 (0.323)	0.172	-0.393 (0.188)	0.036
beta-hydroxyisovaleroylcarnitine	Leucine, Isoleucine and Valine Metabolism	-0.254 (0.292)	0.384	-1.256 (0.394)	0.001	-0.608 (0.234)	0.010
3-methylglutarylcarnitine (2)	Leucine, Isoleucine and Valine Metabolism	-0.171 (0.202)	0.397	-0.532 (0.184)	0.004	-0.368 (0.136)	0.007
N6-carboxyethyllysine	Lysine Metabolism	-0.462 (0.214)	0.031	-0.293 (0.216)	0.174	-0.378 (0.152)	0.013
5-(galactosylhydroxy)-lysine	Lysine Metabolism	-0.778 (0.369)	0.035	-0.400 (0.394)	0.310	-0.601 (0.269)	0.025
6-oxopiperidine-2-carboxylate	Lysine Metabolism	-0.577 (0.351)	0.100	-0.654 (0.400)	0.102	-0.610 (0.264)	0.021
hydroxy-N6,N6,N6-trimethyllysine*	Lysine Metabolism	-0.691 (0.377)	0.067	-0.523 (0.446)	0.241	-0.621 (0.288)	0.031
21-hydroxypregnenolone disulfate	Pregnenolone Steroids	-0.280 (0.238)	0.238	-0.416 (0.246)	0.091	-0.346 (0.171)	0.043

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
uric acid ribonucleoside*	Purine Metabolism, (Hypo)Xanthine/Inosine containing	-0.417 (0.289)	0.149	-0.501 (0.335)	0.135	-0.453 (0.219)	0.039
1-methylhypoxanthine	Purine Metabolism, (Hypo)Xanthine/Inosine containing	-0.596 (0.349)	0.087	-0.637 (0.549)	0.246	-0.608 (0.294)	0.039
5-methyluridine (ribothymidine)	Pyrimidine Metabolism, Uracil containing	0.207 (0.224)	0.356	0.418 (0.203)	0.04	0.323 (0.150)	0.032
riboflavin (vitamin B2)	Riboflavin Metabolism	-0.637 (0.325)	0.05	-0.840 (0.335)	0.012	-0.736 (0.233)	0.002
succinylcarnitine (C4-DC)	TCA Cycle	-0.615 (0.357)	0.085	-0.522 (0.354)	0.14	-0.568 (0.251)	0.024
thiamin (vitamin B1)	Thiamine Metabolism	-0.419 (0.326)	0.198	-0.636 (0.325)	0.05	-0.528 (0.230)	0.022
methylurea	Urea cycle; Arginine and Proline Metabolism	-0.329 (0.213)	0.122	-0.299 (0.247)	0.226	-0.316 (0.161)	0.050
N-monomethylarginine	Urea cycle; Arginine and Proline Metabolism	0.250 (0.226)	0.268	0.394 (0.211)	0.062	0.327 (0.154)	0.034
N,N,N-trimethyl-alanylproline betaine (TMAP)	Urea cycle; Arginine and Proline Metabolism	-0.576 (0.377)	0.126	-0.709 (0.547)	0.195	-0.619 (0.310)	0.046
pyridoxine (vitamin B6)	Vitamin B6 Metabolism	-0.365 (0.232)	0.115	-0.266 (0.217)	0.220	-0.312 (0.158)	0.049
pyridoxate	Vitamin B6 Metabolism	-1.036 (0.501)	0.039	-0.789 (0.491)	0.108	-0.910 (0.351)	0.010
X-12117		-0.667 (0.323)	0.039	-0.694 (0.331)	0.036	-0.680 (0.231)	0.003
X-12708		-0.576 (0.433)	0.183	-0.630 (0.370)	0.088	-0.607 (0.281)	0.031
X-12729		-0.288 (0.120)	0.016	-0.107 (0.165)	0.518	-0.225 (0.097)	0.020
X-12822		-0.848 (0.399)	0.034	-0.306 (0.435)	0.481	-0.600 (0.294)	0.041

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
X-12844		-0.118 (0.091)	0.195	-0.177 (0.075)	0.018	-0.153 (0.058)	0.008
X-15136		-0.577 (0.306)	0.059	-0.379 (0.339)	0.263	-0.488 (0.227)	0.032
X-15523		-0.324 (0.220)	0.140	-0.494 (0.296)	0.095	-0.385 (0.176)	0.029
X-17357		-0.133 (0.107)	0.212	-0.268 (0.106)	0.012	-0.201 (0.075)	0.008
X-18838		-0.324 (0.152)	0.033	-0.380 (0.237)	0.108	-0.341 (0.128)	0.008
X-22152		-0.506 (0.270)	0.061	-0.425 (0.246)	0.083	-0.462 (0.182)	0.011
X-23648		-0.564 (0.356)	0.114	-0.522 (0.306)	0.088	-0.540 (0.232)	0.020
X-23654		-0.431 (0.242)	0.076	-0.932 (0.326)	0.004	-0.609 (0.195)	0.002
X-23918		-0.251 (0.137)	0.067	-0.145 (0.119)	0.225	-0.190 (0.090)	0.034
X-23922		-0.315 (0.229)	0.170	-0.461 (0.179)	0.010	-0.406 (0.141)	0.004
X-24342		-0.130 (0.232)	0.576	-1.098 (0.334)	0.001	-0.445 (0.190)	0.019
X-24468		-0.285 (0.190)	0.134	-0.663 (0.229)	0.004	-0.439 (0.146)	0.003
X-24801		-0.267 (0.253)	0.290	-0.825 (0.275)	0.003	-0.522 (0.186)	0.005
X-25109		-0.142 (0.073)	0.051	-0.131 (0.167)	0.433	-0.140 (0.067)	0.036
X-25422		-0.329 (0.260)	0.204	-1.363 (0.414)	0.001	-0.621 (0.220)	0.005
X-25457		-0.473 (0.176)	0.007	-0.545 (0.275)	0.048	-0.494 (0.148)	0.001

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Table S2.5. Summary statistics of differential metabolites (p<0.05) at time point 2 (postnatal days 23-32) in infants on enteral nutrition

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
N-acetylglucosaminylasparagine	Aminosugar Metabolism	-0.531 (0.413)	0.199	-0.997 (0.599)	0.096	-0.681 (0.340)	0.045
androstenediol (3beta,17beta) disulfate (1)	Androgenic Steroids	-0.410 (0.215)	0.057	-0.352 (0.298)	0.237	-0.390 (0.174)	0.025
androstenediol (3beta,17beta) disulfate (2)	Androgenic Steroids	-0.953 (0.424)	0.025	-0.387 (0.380)	0.309	-0.639 (0.283)	0.024
ascorbate (vitamin C)	Ascorbate and Aldarate Metabolism	-0.025 (0.011)	0.026	0.115 (0.079)	0.145	-0.022 (0.011)	0.046
diisopropanolamine	Chemical	0.491 (0.284)	0.083	0.395 (0.254)	0.120	0.438 (0.189)	0.021
phenylalanylhydroxyproline*	Dipeptide	-0.355 (0.152)	0.019	-0.152 (0.158)	0.338	-0.258 (0.110)	0.019
leucylhydroxyproline*	Dipeptide Derivative	-0.363 (0.144)	0.012	-0.127 (0.170)	0.456	-0.265 (0.110)	0.016
cysteinylglycine disulfide*	Glutathione Metabolism	-0.978 (0.482)	0.043	-0.451 (0.364)	0.216	-0.643 (0.291)	0.027
2,3-dihydroxy-3-methylvalerate	Leucine, Isoleucine and Valine Metabolism	-0.107 (0.068)	0.117	-0.149 (0.103)	0.149	-0.120 (0.057)	0.035
5-(galactosylhydroxy)-lysine	Lysine Metabolism	-0.628 (0.333)	0.060	-0.625 (0.455)	0.170	-0.627 (0.269)	0.020
S-methylcysteine sulfoxide	Methionine, Cysteine, SAM and Taurine Metabolism	-0.231 (0.124)	0.061	-0.160 (0.112)	0.151	-0.192 (0.083)	0.020
mevalonate	Mevalonate Metabolism	-0.631 (0.253)	0.013	-0.586 (0.285)	0.040	-0.611 (0.189)	0.001
glucuronide of C9H10O4 (1)*	Partially Characterized Molecules	-0.355 (0.216)	0.100	-0.250 (0.206)	0.223	-0.300 (0.149)	0.044
4-hydroxyphenylacetate	Phenylalanine Metabolism	-0.085 (0.034)	0.014	-0.006 (0.042)	0.879	-0.053 (0.027)	0.047
5-methylthioadenosine (MTA)	Polyamine Metabolism	0.434 (0.263)	0.099	0.858 (0.601)	0.153	0.502 (0.241)	0.037
isoputreanine	Polyamine Metabolism	-0.632 (0.410)	0.123	-1.099 (0.747)	0.141	-0.740 (0.359)	0.039

Metabolite Name	Sub-pathway	TOLSURF		PROP		Meta	
		Beta (SE)	*p*	Beta (SE)	*p*	Beta (SE)	*p*
sphingosine	Sphingosines	-0.247 (0.126)	0.051	-0.110 (0.148)	0.455	-0.189 (0.096)	0.049
homocitrate	TCA Cycle	-0.445 (0.273)	0.103	-0.309 (0.211)	0.143	-0.360 (0.167)	0.031
tryptophan	Tryptophan Metabolism	-0.931 (0.439)	0.034	-0.335 (0.270)	0.214	-0.499 (0.230)	0.03
kynurenine	Tryptophan Metabolism	-0.619 (0.337)	0.066	-0.342 (0.183)	0.062	-0.405 (0.161)	0.012
3-hydroxykynurenine	Tryptophan Metabolism	-0.615 (0.298)	0.039	-0.284 (0.145)	0.051	-0.348 (0.131)	0.008
4-hydroxyphenylacetoylcarnitine	Tyrosine Metabolism	-0.047 (0.021)	0.022	-0.007 (0.039)	0.85	-0.039 (0.018)	0.034
pro-hydroxy-pro	Urea cycle; Arginine and Proline Metabolism	-0.669 (0.380)	0.078	-0.364 (0.286)	0.203	-0.474 (0.228)	0.038
7-methylxanthine	Xanthine Metabolism	-0.135 (0.092)	0.143	-0.147 (0.104)	0.16	-0.140 (0.069)	0.043
X-12822		-0.748 (0.436)	0.086	-0.973 (0.799)	0.223	-0.799 (0.383)	0.037
X-21364		-0.415 (0.195)	0.034	-0.479 (0.275)	0.082	-0.437 (0.159)	0.006
X-23647		-0.405 (0.370)	0.274	-0.300 (0.163)	0.066	-0.317 (0.149)	0.033
X-23666		-0.428 (0.395)	0.278	-1.085 (0.557)	0.051	-0.648 (0.322)	0.044
X-23787		-0.382 (0.244)	0.117	-0.171 (0.104)	0.102	-0.203 (0.096)	0.034
X-25420		-0.097 (0.049)	0.047	-0.138 (0.114)	0.226	-0.103 (0.045)	0.021

Meta meta-analysis; *Beta* beta-coefficient; *SE* standard error; *p* p-value

Chapter 3: Associations Between Genomic Ancestry Proportions and the Urinary Metabolites of Extremely Premature Infants

Abstract

The human biofluid metabolome is shaped by both environmental and genetic factors, which may include genetic variation that is linked to continental genetic ancestry. Genetic ancestry refers to patterns of genetic variation that differ across haplotypes of historical continental origin, and represents a statistically powerful way to investigate the role of genetics on the metabolome of extremely premature infants. Furthermore, it provides a summation of genetic variation that differs across global populations, genetically admixed populations, and across groups of individuals defined by the social categories of race and ethnicity. We aimed to identify metabolites and metabolic pathways associated with continental genetic ancestry in the urine of extremely premature infants, and to investigate whether the association of genetic ancestry varies between the second and fourth weeks of life. Untargeted metabolomics was performed on 524 urine samples from 262 premature infants (<27 weeks gestational age), collected at two postnatal time points (days 5-15 and 23-32). Genomic ancestry proportions (African, European, Amerindigenous) were inferred via ADMIXTURE using a three-population model across different genetic platforms. Associations between African genomic ancestry proportions and metabolite levels were tested using a linear mixed model followed by metabolite set enrichment analyses. Among 851 urinary metabolites quantified, two involved in histidine metabolism were significantly associated with African genomic ancestry (trans-urocanate, $p = 7.6 \times 10^{-6}$, $q = 6.5 \times 10^{-3}$; cis-urocanate, $p = 1.1 \times 10^{-4}$, $q = 0.04$). In addition, 106 metabolites were associated with African genomic ancestry at $p<0.05$ and were significantly enriched for the Nucleotide super-pathway ($p = 1.7 \times 10^{-7}$, $q = 1.4 \times 10^{-6}$) and depleted in the Lipid super-pathway ($p = 4.3 \times 10^{-4}$, $q = 1.7 \times 10^{-3}$).

Top metabolites associated with African ancestry among infants of maternal Black/African American race/ethnicity were significantly enriched for Histidine metabolism ($p = 1.5x10^{-5}$, $q = 0.001$), and corroborated our findings in the full cohort. Collectively, our results demonstrate that genetic variation linked to African ancestry contributes to variation in the urinary metabolome of extremely premature infants, notably metabolites involved in histidine, nucleotide, and lipid metabolism. In addition, genetic ancestry is an important consideration in metabolomic and potentially clinical studies of extremely premature infants.

Introduction

Differences in metabolite profiles have been reported between term and premature infants, with variation observed even among premature infants.[1,2] This heterogeneity in metabolite levels among premature infants reflects, in part, the rapid physiological changes during postnatal adaptation, developmental immaturity, clinical status, and environmental factors such as feeding modality and clinical care.[3–7] However, one factor that remains largely unexplored is the extent to which genetic variation contributes to this heterogeneity in metabolite levels.

Although extensive evidence supports the contribution of genetic variation to interindividual differences in metabolite profiles in adults,[8–10] evidence in neonates remains limited. Nevertheless, studies in term infants suggest that genetic factors play an important role in shaping neonatal metabolite profiles. A twin study in term neonates estimated high heritability for several metabolites, particularly short-chain acylcarnitines, with heritability estimates ranging from 61-83%.[11] More recently, a genome-wide association study of neonatal metabolites extended these findings by estimating metabolite heritability in unrelated individuals, reporting a mean heritability of 76.2%.[12] While these studies provide a strong foundation for understanding the genetic architecture of metabolites included in newborn metabolic screening, they are limited to single-time-point assessments in term infants and a targeted metabolite panel, leaving unanswered whether these findings extend to extremely premature infants, apply across the metabolome, or change during early postnatal development.

Genomic ancestry, as used throughout this manuscript, refers to ancestry inferred from patterns of genome-wide genetic similarity to reference populations, reflecting aspects of shared population history.[13] These patterns of genetic variation, including differences in allele frequency and linkage disequilibrium, have provided valuable insights into the genetic architecture of complex traits.[14,15] Whether genomic ancestry contributes to downstream molecular phenotypes,

such as metabolites, remains understudied. Nonetheless, a study among admixed adults performed admixture mapping to evaluate the contribution of genomic ancestry to variation in metabolite levels and identified several metabolites associated with genomic ancestry, including lipid polyunsaturated fatty acids and N-acetylated amino acids.[16] However, whether genomic ancestry contributes to metabolite variation in extremely premature infants during this critical developmental period remains unknown.

In this study, we performed metabolome-wide association analyses between metabolite levels and genomic ancestry proportions in 262 extremely premature infants (gestational age <27 weeks) from two clinical cohorts. We performed sub-analyses stratifying infants by maternal race/ethnicity to assess whether associations between genomic ancestry and metabolite levels extended beyond potential social and environmental factors that may correlate with race/ethnicity. We evaluated whether genomic ancestry contributed disproportionately to variation in metabolites from specific metabolic pathways. Additionally, we evaluated whether the contribution of genomic ancestry to variation in metabolite levels differed between the second and fourth weeks of life.

Methods

Study cohorts

A total of 166 infants from the Trial of Late Surfactant for Prevention of Bronchopulmonary Dysplasia (TOLSURF, ClinicalTrials.gov, NCT01022580) and 96 from the Prematurity and Respiratory Outcomes Program (PROP, NCT01435187) with both genomics and urinary metabolomics data were included in our study. Clinical and demographic data collected for both cohorts included gestational age, birthweight, maternal race/ethnicity, administration of corticosteroids and caffeine, feeding modality (parenteral vs enteral nutrition), and respiratory parameters, including respiratory severity score (RSS; FiO_2 × mean airway pressure) and bronchopulmonary dysplasia (BPD) status. Maternal race and ethnicity were self-identified and

collected at the time of infant enrollment as part of the case report form; paternal race/ethnicity was not collected. Maternal race and ethnicity were subsequently combined into a single race/ethnicity variable for analysis, resulting in the categories Non-Hispanic White, Non-Hispanic Black/African American, and Hispanic/Latine White. For analyses and reporting, the Non-Hispanic Black/African American and Hispanic/Latine White categories are referred to as Black/African American and Hispanic/Latine, respectively, throughout this manuscript. Complete clinical and demographic data were available for all included infants. Research protocols for both studies were approved by the institutional review boards of the participating institutions, and a parent of each infant provided written informed consent. The requirement for supplemental oxygen or mechanical ventilation at 36 weeks postmenstrual age was used across both cohorts to diagnose BPD.[17]

TOLSURF was a multicenter, randomized, double-blind, sham-controlled clinical trial that evaluated whether late surfactant administration improved respiratory outcomes in extremely premature infants. Between 2010 and 2013, the trial enrolled 511 infants across 25 centers in the United States. Eligibility criteria included gestational age between 23 and 28 weeks and the requirement for intubation and mechanical ventilation between 7 and 14 days of life. Additional details regarding the study design, enrollment criteria, participant characteristics, and primary outcomes have been reported previously.[17]

PROP was a multicenter prospective observational cohort study designed to evaluate respiratory morbidity and long-term respiratory outcomes in extremely premature infants. Between 2010 and 2013, 835 infants born at <29 weeks of gestational age were enrolled across 8 centers in the United States. Additional information regarding the study design, enrollment criteria, participant characteristics, and study outcomes can be found elsewhere.[18]

Metabolomics profiling

Urine samples were collected from infants at two time points as previously described.[7,19] The first collection was between postnatal days 5-15 (time point 1; mean/SD: 8±2 days) and the second between postnatal days 23-32 (time point 2; mean/SD: 28±2 days). A total of 628 urine samples were obtained from 314 infants who survived to 40 weeks gestational age and had urine samples available at both time points. At time point 2, infants from TOLSURF were categorized as receiving enteral nutrition if they had been off parenteral nutrition for at least four days prior to urine collection, and as receiving parenteral nutrition if they had received it within four days of urine collection. In PROP, at time point 2, infants were categorized as receiving enteral nutrition if they had received full enteral feeds within four days prior to urine collection, and as receiving parenteral nutrition if they had received partial or no enteral feeds within the same period.

Untargeted metabolomics profiling was performed in two batches for TOLSURF and PROP, with anchor samples included to align cross-cohort measurements. Metabolomic assays were performed using Ultra-High-Performance Liquid Chromatography with Tandem Mass Spectrometry (UHPLC-MS/MS) by Metabolon Inc.; additional details are reported elsewhere.[7,19] Briefly, 1,441 metabolites were profiled across the two time points, including 377 (26.2%) and 1,064 (73.8%) metabolites of unknown and known identity, respectively. Metabolites of known identity included amino acids, lipids, carbohydrates, nucleotides, peptides, cofactors and vitamins, and xenobiotics. Each analytical batch included internal standards to adjust for instrument variability. Urine samples were normalized to osmolality. Metabolomics data were preprocessed by Metabolon. Metabolomics data from each cohort at both time points were combined into a single matrix. Metabolites with more than 30% missing values within either cohort were excluded from analysis. For the remaining metabolites, missing values were imputed to the minimum detectable level and data were median scaled. Only metabolites detected at both time points were included in the analyses.

Genomics data

Genotyping in TOLSURF was performed on DNA extracted from tracheal aspirate cells from 454 infants using the Affymetrix Axiom LAT 1 Array (World Array 4), which contained >800,000 variants prior to quality control. Standard quality control was performed as previously described,[20] retaining 409 infants and 795,465 variants. Data were in hg19 coordinates and were lifted over to GRCh38 using the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver); 770,775 of 795,465 variants (96.9%) lifted over successfully.

In PROP, genetic data were available from two platforms. Exome sequencing was performed in two batches at the University of Washington Genome Sciences using the Roche NimbleGen SeqCap EZ Exome Solution v2 capture kit on an Illumina sequencing platform, from saliva samples from 382 infants.[21] Harmonization and standard quality control were performed. Briefly, reads were aligned to hg19 and variant calling was performed jointly across both batches. Following joint variant calling, quality control included recoding of variant calls with sequencing depth (DP) <5 or genotype quality (GQ) ≤25 as missing, prior to standard quality control filters. A separate subset of 240 infants was genotyped using the Illumina Infinium Multi-Ethnic Genotyping Array with custom content (Custom MEGA v2.0) developed by the CARGO lab at the Colorado Center for Personalized Medicine (University of Colorado Anschutz Medical Campus), which contained >1,700,000 variants prior to quality control. Of these, 73 infants had data available from both platforms and were used to assess cross-platform concordance of ancestry estimates (described below).

For the genotyping array data, variant calling and initial quality control were performed using GenomeStudio and PLINK v.1.9. Variants and samples were removed based on 95% call rate, retaining 173 infants and 1,625,562 autosomal variants. Further quality control was performed using PLINK v1.9, with additional variants removed based on Hardy-Weinberg

equilibrium *p*-value $>10^{-6}$. Principal component analysis (PCA) was performed on linkage disequilibrium (LD) pruned single nucleotide polymorphisms (SNPs) ($R^2 > 0.5$ in a 50-SNP window with a shift size of 5 SNPs) to evaluate population structure and identify potential outliers. Subjects with sex discordance and PCA outliers were removed from further analyses. Familial relatedness was assessed using KING to confirm known familial relatedness. A total of 147 infants and 1,439,814 autosomal SNPs were retained. Genetic data from both platforms were lifted over from hg19 to GRCh38 using the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Liftover was successful for 264,428 of 264,644 exome SNPs (99.9%) and 1,332,504 of 1,439,814 genotyping array SNPs (92.5%).

Genomic ancestry proportion estimates

Genomic data were merged separately for each cohort and platform, and with select individuals from the 1000 Genomes Project and Mexican Biobank using *BCFtools,*[22] restricting to autosomal SNP markers that overlapped within each analysis. SNPs in the merged datasets were filtered based on minor allele frequency (MAF < 0.05), and LD pruning was performed ($R^2 > 0.2$ in a 50-SNP window with a shift size of 5 SNPs). The number of SNPs retained per cohort and platform for genomic ancestry proportion inference is summarized in Figure 3.1. Genomic ancestry proportions for each individual were estimated using ADMIXTURE[23] in a supervised analysis assuming three ancestral continental populations of origin (K=3; African, European, and Amerindigenous). These genomic ancestry proportions represent ancestry inferred from patterns of genetic similarity to the predefined reference populations. Unrelated, non-admixed individuals from the 1000 Genomes Project[24] and Mexican Biobank[25] were included as reference populations, representative of African ancestry [Yoruba in Ibadan, Nigeria (YRI), n = 103], European ancestry [Utah residents with Northern and Western European ancestry (CEU), n = 93], and Amerindigenous ancestry [Mexican residents with Amerindigenous ancestry, n = 50].

Concordance between ancestry estimates inferred from exome sequencing vs. genotyping array data was evaluated for each genomic ancestry proportion for infants in PROP using Lin's concordance correlation coefficient (CCC).[26] Overall concordance between exome sequencing and genotyping data was high with negligible systematic bias (CCC = 0.978; C_b = 0.999). Ancestry-specific concordance was consistently high for African and European ancestries, with lower concordance for Amerindigenous genomic ancestry (CCC = 0.838, C_b = 0.818; Supplemental Table S3.1). Ancestry proportions from the exome sequencing platform were used in subsequent analyses for infants with data from both platforms, with genotyping array estimates used only for infants without exome data, given that this subset was smaller. Final ancestry estimates across all infants, cohorts, and platforms were combined into a single matrix for downstream analyses.

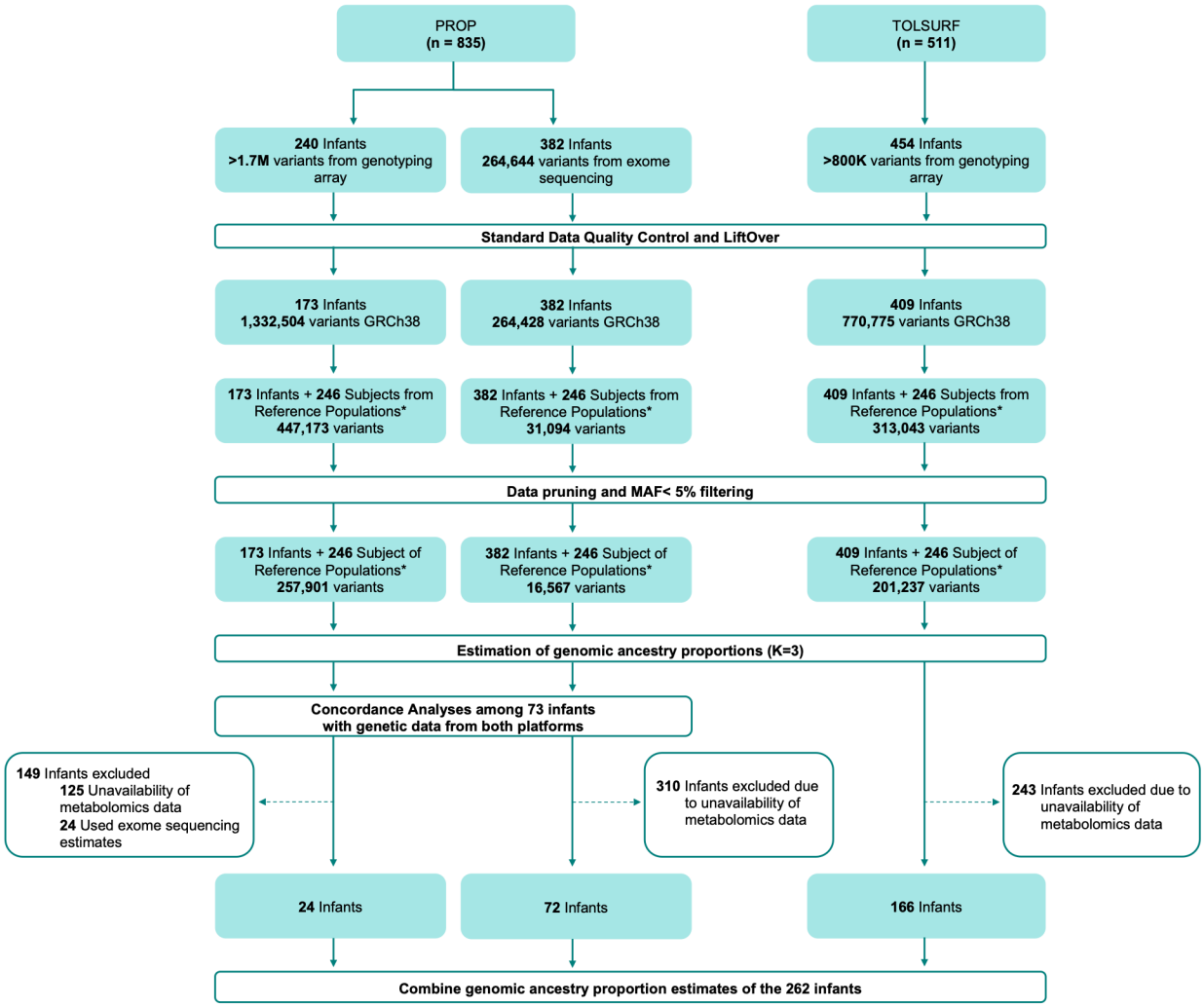

Figure 3.1. Flowchart of infant selection and genomic ancestry estimation workflow.

Overview of infant selection, genetic data processing, genomic ancestry estimation under a three-population ancestry model, concordance analyses, and assembly of the final analytic cohort from the PROP and TOLSURF studies. The final analytic cohort included 262 infants with both genomic ancestry estimates and metabolomics data. *Reference populations included Yoruba in Ibadan, Nigeria (YRI) and Utah residents with Northern and Western (CEU), representative of African and European ancestry, respectively, from the 1000 Genomes Project, and Mexican residents representative of Amerindigenous Ancestry (MEX) from the Mexican Biobank.

Statistical Analyses

Analyses were restricted to infants with both genomics and metabolomics data at both time points (n = 262). A primary linear mixed model was used to evaluate associations between metabolite levels and untransformed proportions of African genomic ancestry, as compared to

European and Amerindigenous proportions combined (non-African genomic ancestry). European and Amerindigenous ancestry components were combined for two reasons. First, as genomic ancestry proportions sum to one, including all ancestry components simultaneously would introduce perfect multicollinearity. Additionally, the Amerindigenous ancestry component showed a sparse distribution (210 [80.2%] of infants have proportions below 5%), with insufficient variation to evaluate it as an independent exposure. Fixed effects included sex, gestational age, birthweight, maternal race/ethnicity, corticosteroid administration, bronchopulmonary dysplasia status, parenteral nutrition status (parenteral vs. enteral nutrition), time point, and cohort. A random intercept for infant ID was included to account for repeated metabolite measures across time points. A secondary linear mixed model was fit among metabolites nominally (p<0.05) associated with African genomic ancestry proportions in the primary model, adding an ancestry-by-time interaction term to evaluate whether these associations varied across time points.

Sub-analyses were performed by stratifying infants by maternal race/ethnicity to further delineate the contribution of African genomic ancestry proportions independent of potential shared environmental effects that may correlate with race/ethnicity. These sub-analyses were restricted to infants of Black or African American and Hispanic/Latine maternal race/ethnicity, as infants of Non-Hispanic White maternal race had uniformly low African genomic ancestry proportions, which did not allow for evaluation of the effect of African genomic ancestry proportions within this group. The primary model without the interaction term was fit separately for each maternal race/ethnicity group, using the same specification as in the primary analysis.

Metabolite set enrichment analysis (MSEA) was performed on metabolites that were associated with untransformed African genomic ancestry proportions at a nominal p<0.05. Analyses were restricted to metabolites of known identity (Metabolon library, as of April 2025), which were annotated to 8 super-pathways and 118 sub-pathways; sub-pathways with fewer than 2 annotated metabolites were excluded. MSEA was performed for each super- and sub-pathway

using Fisher's exact test, with the enrichment ratio calculated as the proportion of nominally significant metabolites relative to the number of insignificant metabolites within a pathway, divided by the proportion of nominally significant metabolites among all metabolites of known identity. MSEA was performed separately for the primary analysis and stratified sub-analyses.

Linear mixed models were fit using the *lmerTest* package, and contrast coefficients and *p*-values were obtained using the *emmeans* package in R (4.4.3). MSEA was performed using a custom Python (3.11.4) script. Multiple comparisons across analyses were accounted for by adjusting *p*-values using the Benjamini-Hochberg method,[27] with *q*-values < 0.05 considered statistically significant.

Table 3.1. Clinical and demographic characteristics of infants from TOLSURF and PROP

	TOLSURF (n = 166)	**PROP (n = 96)**	***p***
Gestational Age (weeks)	25.3 (1.2)	25.8 (1.2)	0.006
Birth weight (g)	722 (172)	790 (144)	0.001
Sex (male)	98 (59.0)	51 (53.1)	0.367
Number of Infants by Maternal Race			<0.0001
Black/African American	58 (35)	43 (45)	
Hispanic/Latine	23 (14)	14 (15)	
Non-Hispanic White	85 (51)	39 (41)	
Average RSS days 7-14	4.2 (1.9)	2.5 (1.4)	<0.0001
Age days sample 1	9.6 (2.3)	6.8 (1.0)	<0.0001
Age days sample 2	27.6 (2.2)	28.1 (1.0)	0.019
BPD at 36 weeks	106 (63.9)	47 (49.0)	0.020
On PN sample 1	166 (100)	96 (100)	1.0
On PN sample 2	105 (63.3)	60 (62.5)	1.0
On corticosteroids sample 1	14 (8.4)	2 (2.1)	0.058
On corticosteroids sample 2	35 (21.2)	26 (27.1)	0.290

p p-value; *RSS* Respiratory severity score; *BPD* bronchopulmonary dysplasia; *PN* parenteral nutrition

Values are represented by the mean (standard error) for continuous variables and the count (percent) for categorical variables. Quantitative measurements were compared using a t-test, and categorical data were compared using a Fisher's exact test.

Results

Study Cohorts

Characteristics of the 166 extremely premature infants from TOLSURF and 96 from PROP included in the analyses are summarized in Table 3.1. Mean gestational age was 25.3 weeks in TOLSURF and 25.8 weeks in PROP, and mean birthweight was 722g and 790g, respectively, with PROP infants having higher gestational age and birthweight on average (p = 0.006, 0.001, respectively). The distribution of maternal race/ethnicity also differed significantly between cohorts (p<0.0001). The majority of TOLSURF infants had mothers who identified as Non-Hispanic White (51%), while the plurality of PROP infants had mothers who identified as Black/African American (45%). Both cohorts included approximately 15% Hispanic/Latine mothers.

Metabolomics and genomics data

Of the 1441 unique metabolites characterized across TOLSURF and PROP through an untargeted metabolomics approach, 851 passed quality control thresholds within each cohort and were detected at both time points. Genomic ancestry proportions are summarized in Figure 3.2. African genomic ancestry proportions varied widely across infants (Figure 3.2A), with clustering patterns reflecting maternal race/ethnicity (Figure 3.2B). The distribution was bimodal, with 95 infants (36.3%) having proportions below 0.5%. As expected, African ancestry proportions varied by maternal race/ethnicity, with all 101 infants of Black/African American maternal race/ethnicity having proportions above 30% (Figure 3.2D). Infants of maternal Non-Hispanic White race/ethnicity (n = 124) had uniformly low African ancestry proportions [117 (94.4%) infants below 5%], while infants of maternal Hispanic/Latine race/ethnicity showed greater variability, with proportions ranging from 0% to 46.4% [24 (64.9%) infants above 5%, Figure 3.2E].

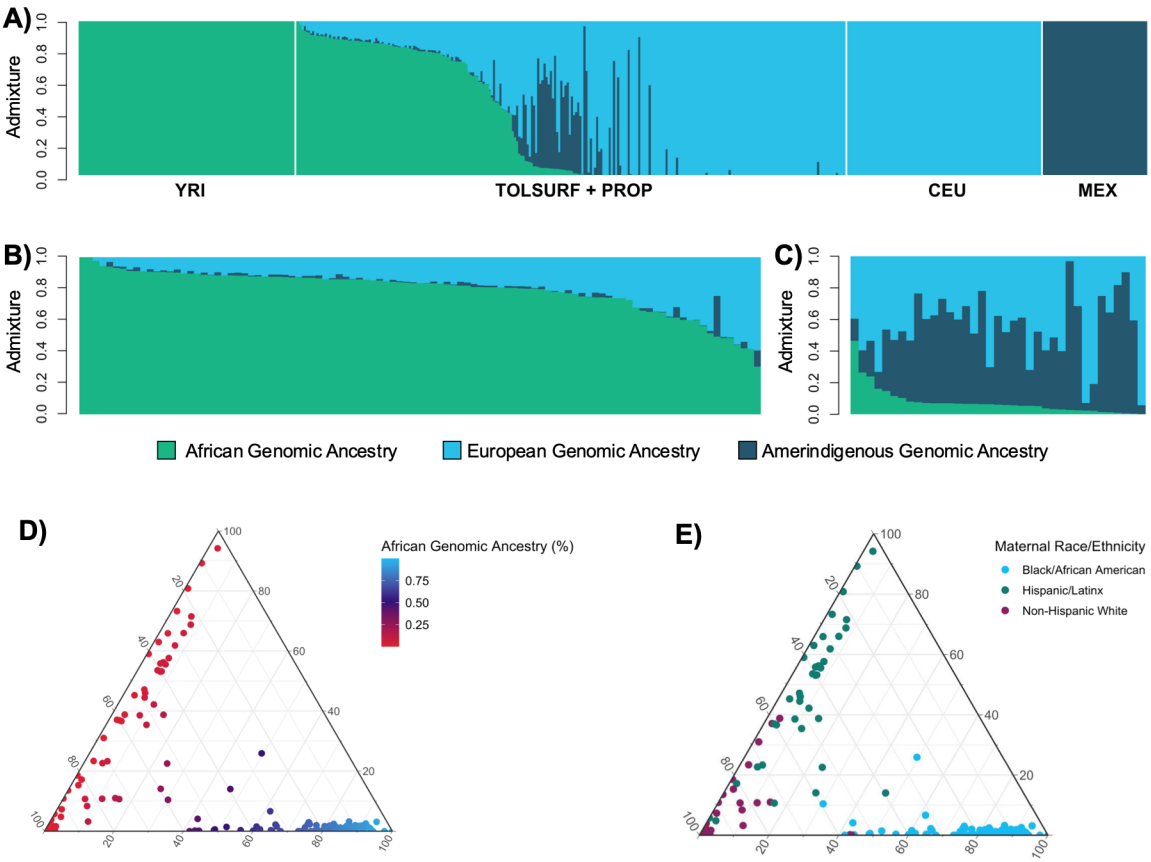

Figure 3.2. Genomic ancestry estimates and composition of the study cohorts.
ADMIXTURE analysis (*K* = 3) was performed using African (represented by YRI), European (represented by CEU), and Amerindigenous (represented by MEX) reference populations together with TOLSURF and PROP participants. Each vertical bar represents one individual, and colors indicate the estimated proportion of African (green), European (light blue), and Amerindigenous (dark blue) genomic ancestry. Individuals within each population are ordered by decreasing African genomic ancestry. (A) Individual ancestry estimates for reference populations and study participants. (B) Individual ancestry estimates for infants of self-reported Black/African American maternal race/ethnicity. (C) Individual ancestry estimates for infants of self-reported Hispanic/Latine maternal race/ethnicity. (D) Ternary plot of individual genomic ancestry proportions, with points colored according to African genomic ancestry. (E) Ternary plot of individual genomic ancestry proportions, with points colored according to self-reported maternal race/ethnicity.

Metabolome-Wide Associations with African Genomic Ancestry

Associations between African genomic ancestry proportions and urinary metabolite levels were evaluated across all infants and time points, with sub-analyses stratified by maternal race/ethnicity. Of the 851 metabolites evaluated, 106 were associated with African genomic

ancestry proportions at a nominal $p<0.05$ (top metabolites), representing a 2.5-fold enrichment over what would be expected by chance (43 expected at $p<0.05$) (Figure 3.3, Supplemental Table S3.2). Of these top metabolites, 79 were of known identity. Three of the 106 metabolites were statistically significant at $q < 0.05$, two of known identity within the histidine metabolism pathway, and one of unknown identity (Table 3.2). Higher levels of trans-urocanate and cis-urocanate (histidine metabolism) were associated with lower proportions of African genomic ancestry, while higher levels of the metabolite of unknown identity (X-12729) were associated with higher proportions. Notably, 96 of the 106 top associated metabolites (91%) were positively associated with African genomic ancestry proportions (odds ratio = 0.34, $p = 0.0006$), a directional pattern also observed among infants of Hispanic/Latine (odds ratio = 0.09, $p = 5.9\text{x}10^{-8}$) but not Black/African American (odds ratio = 0.71, $p = 0.43$) maternal race/ethnicity.

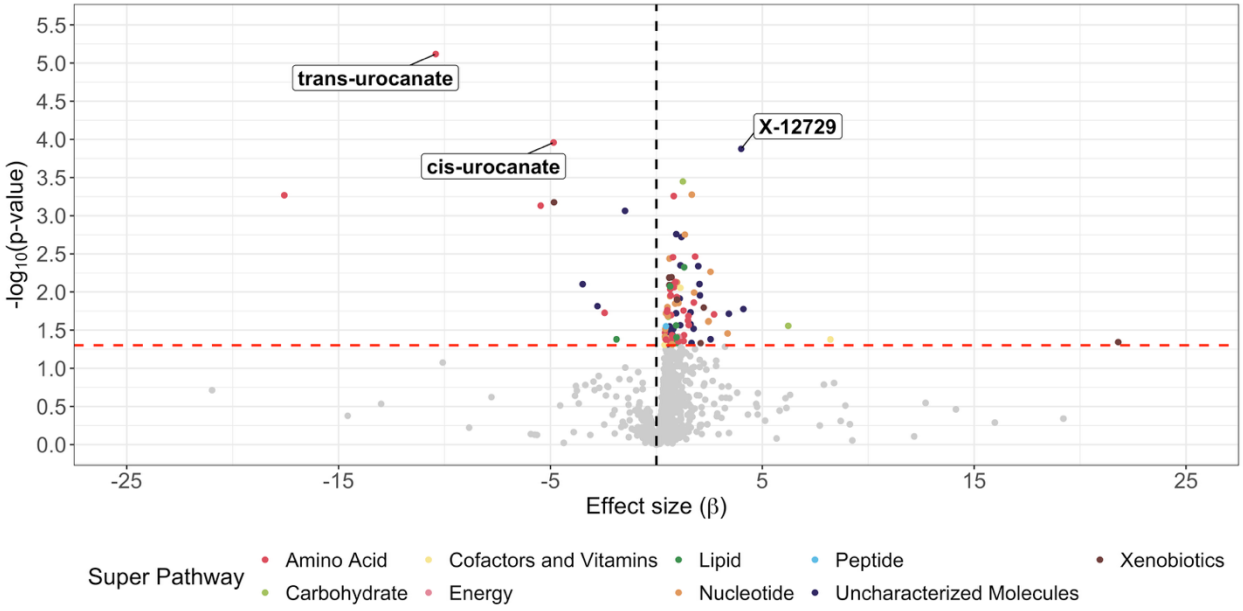

Figure 3.3. Metabolome-wide association analysis of African genomic ancestry.
Volcano plot summarizing linear mixed-effects model results for the association of African genomic ancestry with individual metabolite levels. Each point represents one metabolite, with the x-axis indicating the estimated regression coefficient (β) for African genomic ancestry and the y-axis showing the corresponding $-\log_{10}(p\text{-value})$. The red dashed horizontal line denotes the nominal significance threshold ($p = 0.05$), and the vertical dashed line indicates β = 0. Metabolites are colored by super pathway. Metabolites meeting the false discovery rate threshold (q-value < 0.05) are labeled.

Table 3.2. Metabolites significantly associated with African genomic ancestry proportions

Metabolite	Sub-pathway	Super-pathway	Beta	SE	*p*	*q*
cis-urocanate	Histidine Metabolism	Amino Acid	-4.853	1.235	0.00011	0.038
trans-urocanate	Histidine Metabolism	Amino Acid	-10.416	2.278	0.00001	0.006
X-12729	Uncharacterized	Uncharacterized	4.002	1.031	0.00013	0.038

Beta beta-coefficient; *SE* standard error; *p* p-value; *q* q-value

Ancestry-by-time point interactions were evaluated among the 106 top metabolites associated with African genomic ancestry proportions, with 4 metabolites reaching nominal significance for the interaction term ($p<0.05$), none of which were significant after multiple testing correction (Supplemental Table S3.3). Of the 4 nominally associated metabolites, three were of known identity, including one amino acid (dimethylguanidino valeric acid (DMGV)*) and two xenobiotic (monobutyl phthalate acyl-beta-glucuronide and 4-hydroxymandelate) metabolites. The nominal interaction reflected time point-specific differences in the strength of ancestry associations, with stronger associations at time point 1 for monobutyl phthalate acyl-beta-glucuronide and at time point 2 for 4-hydroxymandelate, dimethylguanidino valeric acid (DMGV)*, and the metabolite of unknown identity (X-24411) (Figure 3.4).

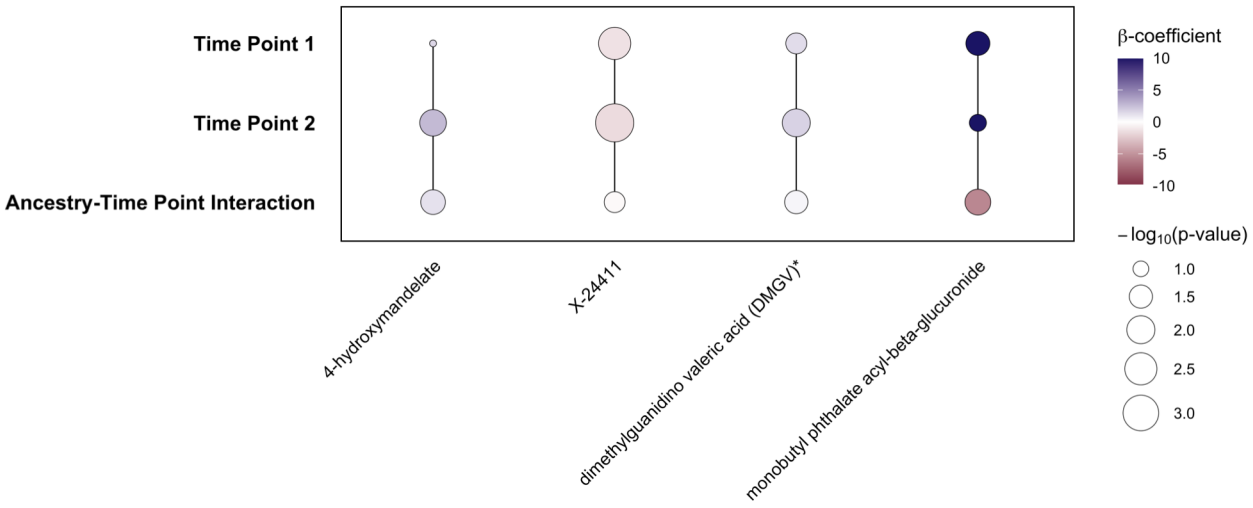

Figure 3.4. Metabolites associated at p<0.05 with the ancestry-by-time point interaction
Bubble plot summarizing results from linear mixed-effects models evaluating the interaction between African genomic ancestry and time point on metabolite levels. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) Bubble color represents the estimated regression coefficient (β), with blue indicating positive coefficients and red indicating negative coefficients. Bubble size is proportional to the statistical significance $-\log_{10}(p\text{-value})$. Rows correspond to the estimated effects of African genomic ancestry at Time Point 1, African genomic ancestry at Time Point 2, and the ancestry-by-time point interaction. Only metabolites with a nominally significant ancestry-by-time point interaction (*p-value* < 0.05) are shown. No interactions remained significant after false discovery rate correction (*q*-value < 0.05).

Table 3.3. Summary table of nominally associated metabolites across all analyses

		All Infants			Black/African American			Hispanic/Latine		
Metabolite	**Super-pathway**	**Beta**	***p***	***q***	**Beta**	***p***	***q***	**Beta**	***p***	***q***
arabinose	Carbohydrate	1.251 (0.346)	0.0004	0.0672	1.293 (0.496)	0.011	0.570	3.435 (1.007)	0.0018	0.2419
glycerophosphoethanolamine	Lipid	1.312 (0.460)	0.0047	0.2121	1.450 (0.675)	0.034	0.662	3.189 (1.002)	0.0033	0.2521
glycerophosphoserine*	Lipid	0.649 (0.244)	0.0084	0.2420	0.729 (0.352)	0.041	0.665	1.692 (0.629)	0.0112	0.4472
guanosine-3',5'-cyclic monophosphate (cGMP)	Nucleotide	1.672 (0.476)	0.0005	0.0672	1.492 (0.574)	0.011	0.570	4.927 (1.490)	0.0023	0.2419
uric acid ribonucleoside*	Nucleotide	0.736 (0.371)	0.0483	0.3941	0.949 (0.476)	0.049	0.665	1.892 (0.772)	0.0200	0.5092
X-12729	Uncharacterized	4.002 (1.031)	0.0001	0.0377	3.864 (1.736)	0.028	0.658	6.335 (2.635)	0.0223	0.5096

Beta beta-coefficient; *SE* standard error; *p* p-value; *q* q-value

Sub-analyses stratified by maternal race/ethnicity were performed among infants of Black/African American or Hispanic/Latine maternal race/ethnicity to further examine potential confounding given the high correlation between genetic ancestry and race/ethnicity. Of the 106 top metabolites associated with African genomic ancestry proportions across all infants, 45 were also nominally significant in the sub-analysis of infants of Black/African American maternal race/ethnicity, with consistent directions of effect across analyses (Supplemental Table S3.4). These included the 3 statistically significant metabolites (trans-urocanate, cis-urocanate, and X-12729) identified in the primary analysis. A similar trend was observed among infants of Hispanic/Latine maternal race/ethnicity, where 22 of the 106 top ancestry-associated metabolites were nominally significant in the sub-analysis, with a consistent directions of effect across analyses, including X-12729 (Supplemental Table S3.5). Notably, six metabolites reached nominal significance across both maternal racial/ethnic groups and the primary analysis, with consistent directions of effects across all three analyses (Table 3.3); of these, only X-12729 was statistically significant in the primary analysis. None of the 59 and 60 nominally significant metabolites identified in the sub-analyses of infants of Black/African American and Hispanic/Latine maternal race/ethnicity, respectively, reached the threshold for statistical significance (Supplemental Table S3.6-S3.7).

Metabolite Set Enrichment Analyses

Metabolite set enrichment analyses were performed on metabolites associated with ancestry at $p < 0.05$ (top metabolites) using the super- and sub-pathway annotations provided by Metabolon. Enrichment analysis included 8 super-pathways and 71 sub-pathways with ≥2 metabolites. The Nucleotides super-pathway (enrichment ratio = 3.2, $p = 1.7 \times 10^{-7}$, $q < 0.0001$) was enriched among top metabolites associated with African genomic ancestry proportions (Figure 3.5), a finding consistent with the stratified analysis among infants of Black/African American maternal race/ethnicity (enrichment ratio = 3.5, $p = 6.2 \times 10^{-6}$, $q < 0.0001$). The Lipids super

pathway was depleted among top metabolites associated with proportions of African genomic ancestry (enrichment ratio = 0.34, $p = 4.3 \times 10^{-4}$, $q = 0.0017$), but was not significant in sub-analyses of infants stratified by maternal race/ethnicity. Among sub-pathways, metabolites involved in pyrimidine metabolism, uracil containing were significantly enriched among top ancestry-associated metabolites in the primary analysis (enrichment ratio = 3.9, $p = 6.6 \times 10^{-4}$, $q = 0.047$), while metabolites involved in histidine metabolism were significantly enriched in the sub-analysis of infants of Black/African American maternal race/ethnicity (enrichment ratio = 5.1, $p = 1.5 \times 10^{-5}$, $q = 0.0010$). There was no significant enrichment of top metabolites in any super- or sub-pathway in the sub-analysis of infants of Hispanic/Latine maternal race/ethnicity.

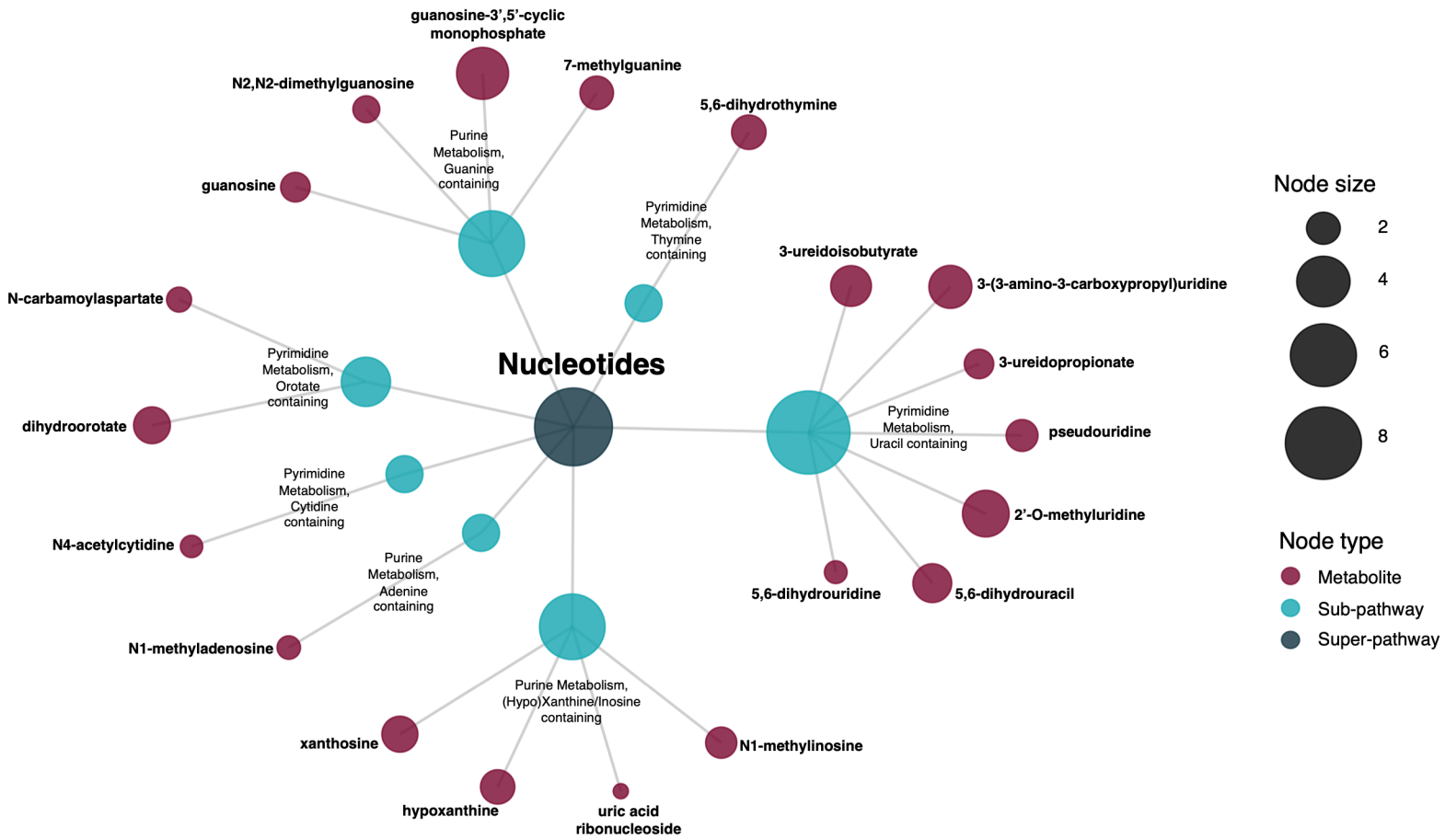

Figure 3.5. Pathway network of metabolites associated with African genomic ancestry within the enriched Nucleotide super-pathway.

Network illustrating the hierarchical organization of metabolites nominally associated with African genomic ancestry (p<0.05) within sub-pathways of the Nucleotide super-pathway, which was significantly enriched by metabolite set enrichment analysis (MSEA). Metabolites (purple nodes) are connected to their corresponding sub-pathways (teal nodes), which are connected to their respective super-pathway (dark teal node). Node size is proportional to the number of connected nodes (degree).

Discussion

In this study, we evaluated the association of continental genomic ancestry with variation in metabolite levels in the urine of extremely premature infants within the first month of life in the Neonatal Intensive Care Unit (NICU). Comparisons of continental ancestry across genetically diverse admixed populations of individuals are a statistically powerful approach to quantify the role of genetic variation that differs in frequency across continental populations on complex traits.[15,28] However, this approach has yet to be applied to the urinary metabolome of infants exposed to extreme prematurity and early life clinical interventions. Overall, we found that African as compared to non-African genomic ancestry proportions were associated with variation in a number of individual metabolites, and top associated metabolites were enriched for specific metabolic pathways as discussed in more detail below. These metabolites and metabolic pathways, therefore, represent those more likely influenced by genetic variation that varies in frequency across infant populations, as compared to variation in environmental, social, and clinical interventions alone. Lastly, we found that the association of African ancestry with the infant metabolome was consistent across two time points within the first month of life, suggesting the contribution of ancestry-linked genetic variation may be relatively stable during early interventions received during the NICU stay.

In our examination of individual metabolites, trans- and cis-urocanate isomers of urocanic acid, key histidine catabolites, showed highly significant associations with African genomic ancestry proportions, with higher levels observed in infants with lower African genomic ancestry proportions. We also observed an enrichment of top metabolites with functions in histidine metabolism among infants of Black/African American maternal race/ethnicity, suggesting our findings are unlikely due to residual confounding from non-genetic factors that correlate with race. The role of genetics in histidine metabolism, including the levels of urocanic acid isomers in the premature infant, remains poorly understood; however, functional investigations suggest a role in

immune function and inflammatory signaling in the skin.[29,30] More specifically, histidine metabolites are involved in immune regulation, oxidative stress responses, inflammatory signaling, and neurotransmitter signaling,[31] all of which have been implicated in adverse outcomes of prematurity.[32,33] Moreover, lower histidine levels in dried blood spots and exhaled breath condensates have been reported among infants who develop bronchopulmonary dysplasia, a common developmental lung disease in premature infants, while higher levels have been observed in tracheal aspirates in the same population.[34–36]

Previous studies in adult populations have reported that levels of several histidine metabolites are influenced by genetic variation, specifically variants in the *HAL* gene, which encodes histidine ammonia-lyase, the enzyme that catalyzes the conversion of histidine to trans-urocanate.[37–40] Notably, a study of European adults found that rare variants in *HAL* were associated with variation in trans-urocanate levels in urine.[39] Similarly, levels of trans-urocanate have been associated with an intronic *HAL* variant across multiple racial/ethnic groups, with the strongest associations reported in African Americans and Hispanic individuals,[40] which suggests a potential relevance of *HAL* genetic variation beyond factors that may correlate with race/ethnicity. Thus, taken together, our findings provide further support that genetic variation contributes to variation in histidine metabolism, in particular variation that is linked to African genomic ancestry. Future studies aimed at identifying ancestry-related differences in histidine metabolism should therefore consider genetic variation that differs across continental haplotypes, including investigations into the specific role of genetic variation at individual loci in regulating metabolites in this pathway.

We also observed an enrichment of top ancestry-associated metabolites within the Nucleotide super-pathway, and the pyrimidine metabolic sub-pathway (uracil containing) across all infants. Nucleotide metabolism comprises metabolites involved in the synthesis, interconversion, salvage, and degradation of nucleotides, the building blocks of RNA and DNA,

which are essential for cellular functions and energy transfer.[41] In extremely premature infants, variation in the Nucleotide super-pathway has been previously reported in the context of feeding modality (parenteral vs enteral nutrition) associations,[7] and between premature as compared to term infants.[42] A genetic contribution to levels of nucleotide metabolites has been previously reported in a multi-ethnic population of adults, with small effect sizes,[40] consistent with the pathway-level enrichment in our analyses. Notably, the same study reports that variants across several genes contribute to some nucleotide metabolite variation, with consistently small effect sizes, though not all variants were significant across all racial/ethnic groups evaluated, and differences were reported in pediatric populations.[36] Therefore, while a pleiotropic genetic contribution to several nucleotide metabolites has been reported, our observed association with African genomic ancestry may capture, in part, a collection of causal genetic variation summed across the genome. Further studies, including admixture mapping, should be performed to disentangle the genetic contributions to these associations.

In our study, we identified a depletion of top metabolites in the Lipid super-pathway. Although fewer lipid metabolites were among top ancestry-associated metabolites with African genomic ancestry proportions than expected by chance, multiple lipid metabolites have previously been associated with genetic variation in adult populations,[16,40,43–45] representing the super-pathway with the most metabolite-variant associations in a previous study.[40] Of particular relevance to our analyses, a previous study that performed admixture mapping in a population of Hispanic individuals with varying proportions of African genomic ancestry identified an association between lipid polyunsaturated fatty acids and several loci mapped to Amerindigenous ancestry, while associations with African locus-specific ancestry were sparse.[16] Lipid metabolism has been extensively studied in premature infants, having been identified as dysregulated in this population and in mothers who give preterm birth.[46–49] Lipid metabolites play a unique role in energy metabolism, cell structure, and surfactant production, biological processes particularly relevant to

premature infants.[50] These metabolites have also been associated with several diseases particular to this population, such as bronchopulmonary dysplasia[51] and necrotizing enterocolitis.[52] Collectively, this highlights the importance of lipid metabolism in extremely premature infants and the heterogeneous nature of its dysregulation in this population. Taken together, the observed depletion may reflect several possibilities: heterogeneity in lipid dysregulation primarily driven by non-genetic factors such as clinical status and feeding modality, limiting power to detect small ancestry-linked genetic effects; a stronger genetic effect concentrated at a smaller number of loci or genes; genetic variation at similar frequencies across continental ancestries represented in our analyses; or a true lack of contribution to metabolite variation.

We further evaluated whether the association between metabolites and African ancestry proportions varied within the first month of life in extremely premature infants, to assess whether the effects of ancestry-linked genetic variation differed between the 1st and 4th weeks of life. This represents a critical time point in the clinical course of an infant born extremely premature, and thus, we hypothesized that the role of genetic variation may change in magnitude or direction under increasing variability in clinical interventions such as the transition to enteral feeding. Previous studies in adults have highlighted the interplay between genetics and the environment in shaping the metabolome, suggesting that while genetically regulated metabolite levels are temporally stable, factors such as age, diet, and body weight could contribute to inter-individual variation in metabolite levels, changing the relative contribution of genetics to metabolite variation.[53,54] However, we found no significant interaction between African genomic ancestry proportions and metabolite levels across time points for any of the metabolites examined after multiple testing correction. This may indicate a consistent genetic effect on ancestry-associated metabolite levels over time, or a lack of statistical power. Regardless, interactions for several

metabolites were at a nominal level of significance and warrant further investigation, particularly in the context of heterogeneity in clinical trajectories among extremely premature infants.

Our study has several strengths. First, we evaluated the contribution of African genomic ancestry to variation in urinary metabolites using a multi-cohort dataset of diverse extremely premature infants, increasing the generalizability of these findings. Second, the use of an untargeted metabolomics approach enabled agnostic discovery across the urinary metabolome, without prior assumptions about which metabolites or pathways may be influenced by African genomic ancestry proportions. Third, we characterized over 800 individual metabolites at two time points during the first month of life, allowing us to comprehensively profile the urinary metabolome and assess whether ancestry-metabolite associations vary during a critical developmental window. Fourth, we performed stratified analyses by maternal race/ethnicity to provide further assurance that our findings are not confounded by the correlation between genetic ancestry and race/ethnicity. Fifth, by reporting associations with metabolites of unknown identity, our findings may extend beyond the current study as these metabolites become fully characterized, potentially uncovering additional ancestry-metabolite associations of biological relevance. Lastly, while differences in clinical features across cohorts could introduce heterogeneity into our analyses, the associations identified in the cross-cohort analyses are potentially less likely to reflect clinical features specific to either cohort.

There are also important limitations to our study that must be acknowledged. First, combining ancestry estimates from different genetic platforms may have introduced systematic biases that our concordance analyses may not have fully captured. Nonetheless, cross-platform differences were mainly observed in Amerindigenous ancestry estimates, with discordance primarily driven by European ancestry estimates that we combined into a single category of non-African genomic ancestry. Second, our sample size may have limited our statistical power to detect ancestry-metabolite associations and ancestry-by-time point interactions of smaller effect.

This limitation also extends to sub-analyses, particularly among infants of Hispanic/Latine maternal race/ethnicity, where the small sample size and narrow distribution of African genomic ancestry proportions further constrained our analyses. Third, maternal race/ethnicity was categorized using broad, self-reported racial/ethnic groups. These categories may not fully capture the social, environmental, and cultural heterogeneity within groups, potentially limiting our ability to more fully distinguish the independent effect of African genomic ancestry from social and environmental factors that may correlate with race/ethnicity. Additionally, paternal race/ethnicity was not collected, preventing evaluation of parental race/ethnicity from both parents. Fourth, while genomic ancestry proportions are highly informative, they represent the summation of genetic variation across the genome rather than locus-specific effects, and thus, the associations identified reflect cumulative effects across loci. Finally, the urine samples used for metabolomics were stored for up to 10 years prior to characterization, potentially causing oxidation or degradation of key metabolites. Therefore, future investigations in larger independent cohorts of extremely premature infants are required to ensure the generalizability of our findings.

In conclusion, African genomic ancestry was associated with variation in the levels of individual urinary metabolites among extremely premature infants. These associations extended beyond individual metabolites to broader metabolic pathways, suggesting that genetic ancestry contributes to metabolic variation early after birth. Future research is needed to evaluate the role of ancestry-linked genetic variation in the metabolome of extremely premature infants, including locus-specific admixture mapping to identify the specific causal variation involved, and to evaluate the potential clinical implications.

References

1. Wilson, K. *et al.* Metabolomics of prematurity: analysis of patterns of amino acids, enzymes, and endocrine markers by categories of gestational age. *Pediatr Res* **75**, 367–373 (2014).

2. Abdolahpour, S., Gholami, M., Mohsenipour, R. & Abbasi, F. Metabolic subtypes and biomarkers in preterm and term neonates via targeted screening. *Sci Rep* https://doi.org/10.1038/s41598-026-50955-8 (2026) doi:10.1038/s41598-026-50955-8.

3. Nilsson, A. K. *et al.* Longitudinal Serum Metabolomics in Extremely Premature Infants: Relationships With Gestational Age, Nutrition, and Morbidities. *Front Neurosci* **16**, 830884 (2022).

4. Meneghelli, M. *et al.* Metabolomic Patterns at Birth of Preterm Newborns with Extrauterine Growth Restriction: Towards Putative Markers of Nutritional Status. *Metabolites* **15**, 518 (2025).

5. Moltu, S. J. *et al.* Urinary Metabolite Profiles in Premature Infants Show Early Postnatal Metabolic Adaptation and Maturation. *Nutrients* **6**, 1913–1930 (2014).

6. Bibiloni, P. *et al.* Neonatal Urine Metabolic Signature Reflects Multisystemic Adaptations Linked to Preterm Birth. *International Journal of Molecular Sciences* **26**, (2025).

7. Guardado, M. *et al.* The Urinary Metabolomic Fingerprint in Extremely Preterm Infants on Total Parenteral Nutrition vs. Enteral Feeds. *Metabolites* **13**, 971 (2023).

8. Wang, N. *et al.* Genetic architecture and analysis practices of circulating metabolites in the NHLBI Trans-Omics for Precision Medicine Program. *The American Journal of Human Genetics* **112**, 2720–2738 (2025).

9. Kastenmüller, G., Raffler, J., Gieger, C. & Suhre, K. Genetics of human metabolism: an update. *Hum Mol Genet* **24**, R93–R101 (2015).

10. Hagenbeek, F. A. *et al.* Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun* **11**, 39 (2020).

11. Alul, F. Y. *et al.* The heritability of metabolic profiles in newborn twins. *Heredity* **110**, 253–258 (2013).

12. He, Q. *et al.* A genome-wide association study of neonatal metabolites. *Cell Genomics* **4**, (2024).

13. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research *et al. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. 26902 (National Academies Press, Washington, D.C., 2023). doi:10.17226/26902.

14. Spear, M. L. *et al.* Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. *Elife* **9**, e56029 (2020).

15. Lin, M., Park, D. S., Zaitlen, N. A., Henn, B. M. & Gignoux, C. R. Admixed Populations Improve Power for Variant Discovery and Portability in Genome-Wide Association Studies. *Front Genet* **12**, 673167 (2021).

16. Reynolds, K. M. *et al.* Ancestry-driven metabolite variation provides insights into disease states in admixed populations. *Genome Med* **15**, 52 (2023).

17. Ballard, R. A. *et al.* Randomized Trial of Late Surfactant Treatment in Ventilated Preterm Infants Receiving inhaled Nitric Oxide. *J Pediatr* **168**, 23-29.e4 (2016).

18. Pryhuber, G. S. *et al.* Prematurity and respiratory outcomes program (PROP): study protocol of a prospective multicenter study of respiratory outcomes of preterm infants in the United States. *BMC Pediatr* **15**, 37 (2015).

19. Torgerson, D. *et al.* The hydrocortisone-responsive urinary metabolome of premature infants. *Pediatr Res* **94**, 1317–1326 (2023).

20. Torgerson, D. G. *et al.* Ancestry and genetic associations with bronchopulmonary dysplasia in preterm infants. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L858–L869 (2018).

21. Hamvas, A. *et al.* Exome sequencing identifies gene variants and networks associated with extreme respiratory outcomes following preterm birth. *BMC Genetics* **19**, 94 (2018).

22. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

23. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).

24. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

25. Sohail, M. *et al.* Mexican Biobank advances population and medical genomics of diverse ancestries. *Nature* **622**, 775–783 (2023).

26. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. https://www.scienceopen.com/document?vid=e3cfffed-a777-439b-b244-b4acbaf7f0c7 (2005).

27. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

28. Shriner, D. Overview of Admixture Mapping. https://doi.org/10.1002/cpz1.677 doi:10.1002/cpz1.677.

29. Hug, D. H., Dunkerson, D. D. & Hunter, J. K. The degradation of L-histidine and trans- and cis-urocanic acid by bacteria from skin and the role of bacterial cis-urocanic acid isomerase. *J Photochem Photobiol B* **50**, 66–73 (1999).

30. Hart, P. H. & Norval, M. The Multiple Roles of Urocanic Acid in Health and Disease. *Journal of Investigative Dermatology* **141**, 496–502 (2021).

31. Moro, J., Tomé, D., Schmidely, P., Demersay, T.-C. & Azzout-Marniche, D. Histidine: A Systematic Review on Metabolism and Physiological Effects in Human and Different Animal Species. *Nutrients* **12**, (2020).

32. Wu, T.-J. *et al.* Role of Myeloperoxidase, Oxidative Stress, and Inflammation in Bronchopulmonary Dysplasia. *Antioxidants* **13**, (2024).

33. Martini, S. *et al.* Antenatal and Postnatal Sequelae of Oxidative Stress in Preterm Infants: A Narrative Review Targeting Pathophysiological Mechanisms. *Antioxidants* **12**, (2023).

34. Course, C. W. *et al.* Evidence of abnormality in glutathione metabolism in the airways of preterm born children with a history of bronchopulmonary dysplasia. *Sci Rep* **13**, 19465 (2023).

35. Guo, Y. *et al.* Analysis of variable metabolites in preterm infants with bronchopulmonary dysplasia: a systematic review and meta-analysis. *Ital J Pediatr* **50**, 246 (2024).

36. Piersigilli, F. *et al.* Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics* **15**, 20 (2019).

37. Rhee, E. P. *et al.* An exome array study of the plasma metabolome. *Nat Commun* **7**, 12360 (2016).

38. Yu, B. *et al.* Association of Rare Loss-Of-Function Alleles in HAL, Serum Histidine. *Circulation: Cardiovascular Genetics* **8**, 351–355 (2015).

39. Cheng, Y. *et al.* Rare genetic variants affecting urine metabolite levels link population variation to inborn errors of metabolism. *Nat Commun* **12**, 964 (2021).

40. Feofanova, E. V. *et al.* Whole-Genome Sequencing Analysis of Human Metabolome in Multi-Ethnic Populations. *Nat Commun* **14**, 3111 (2023).

41. Nucleotide Metabolism - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/nucleotide-metabolism.

42. Diaz, S. O. *et al.* Newborn Urinary Metabolic Signatures of Prematurity and Other Disorders: A Case Control Study. *J. Proteome Res.* **15**, 311–325 (2016).

43. Matey-Hernandez, M. L. *et al.* Genetic and microbiome influence on lipid metabolism and dyslipidemia. *Physiol Genomics* **50**, 117–126 (2018).

44. Dron, J. S. & Hegele, R. A. Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* **4**, 130–141 (2016).

45. Gallois, A. *et al.* A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *Nat Commun* **10**, 4788 (2019).

46. Aung, M. T. *et al.* Maternal lipidomic signatures in relation to spontaneous preterm birth and large-for-gestational age neonates. *Sci Rep* **11**, 8115 (2021).

47. La Frano, M. R. *et al.* Umbilical cord blood metabolomics reveal distinct signatures of dyslipidemia prior to bronchopulmonary dysplasia and pulmonary hypertension. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L870–L881 (2018).

48. Goss, K. C. *et al.* Postnatal adaptations of phosphatidylcholine metabolism in extremely preterm infants: implications for choline and PUFA metabolism. *The American Journal of Clinical Nutrition* **112**, 1438–1447 (2020).

49. Qin, X., Ai, F., Zhou, Q., Zhang, Y. & Yan, X. Pre-eclampsia, gestational hypertension, and lipid levels during pregnancy: a systematic review and meta-analysis. *Arch Gynecol Obstet* **312**, 385–402 (2025).

50. Sun, T., Yu, H., Li, D., Zhang, H. & Fu, J. Emerging role of metabolic reprogramming in hyperoxia-associated neonatal diseases. *Redox Biology* **66**, 102865 (2023).

51. Piersigilli, F. & Bhandari, V. Metabolomics of bronchopulmonary dysplasia. *Clinica Chimica Acta* **500**, 109–114 (2020).

52. Burge, K., Vieira, F., Eckert, J. & Chaaban, H. Lipid Composition, Digestion, and Absorption Differences among Neonatal Feeding Strategies: Potential Implications for Intestinal Inflammation in Preterm Infants. *Nutrients* **13**, 550 (2021).

53. Chen, L. *et al.* Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat Med* **28**, 2333–2343 (2022).

54. Wang, J. *et al.* Longitudinal analysis of genetic and environmental interplay in human metabolic profiles and the implication for metabolic health. *Genome Med* **17**, 68 (2025).

Supplementary Material

Supplementary Tables

Table S3.1. Concordance between exome sequencing– and genotyping array–derived ancestry estimates across ancestry components

Ancestry Component	Pearson's r*	Bias Correction Factor (Cb)	Lin's CCC (95% CI)
African	0.997	0.986	0.983 (0.976-0.988)
European	0.981	0.976	0.957 (0.938-0.971)
Amerindigenous	0.917	0.971	0.890 (0.838-0.926)

CCC concordance correlation coefficient; *CI* confidence interval; *Cb* bias correction factor

*Pearson's r derived as CCC/Cb

Table S3.2 Summary statistics for metabolites nominally ($p<0.05$) associated with African genomic ancestry proportions

Available online as a separate file.

Table S3.3 Summary statistics for metabolites evaluated in ancestry-by-time point interaction analyses

Available online as a separate file.

Table S3.4. Summary statistics for metabolites associated with African genomic ancestry proportions (p<0.05) in infants of Black/African American maternal race/ethnicity, restricted to those also nominally significant in the primary analysis

Metabolite	Sub-pathway	Super-pathway	Beta (SE)	*p*	*q*
1-methyl-4-imidazoleacetate	Histidine Metabolism	Amino Acid	0.626 (0.291)	0.034	0.662
1-methylhistidine	Histidine Metabolism	Amino Acid	0.685 (0.311)	0.030	0.658
5-(galactosylhydroxy)-lysine	Lysine Metabolism	Amino Acid	0.869 (0.393)	0.030	0.658
5-hydroxyindoleacetate	Tryptophan Metabolism	Amino Acid	0.946 (0.432)	0.031	0.662
carnosine	Histidine Metabolism	Amino Acid	2.946 (1.008)	0.004	0.529
cis-urocanate	Histidine Metabolism	Amino Acid	-6.720 (1.967)	0.001	0.402
diacetylspermidine*	Polyamine Metabolism	Amino Acid	0.967 (0.427)	0.026	0.658
fructosyllysine	Lysine Metabolism	Amino Acid	-3.358 (1.454)	0.023	0.658
homovanillate (HVA)	Tyrosine Metabolism	Amino Acid	-5.761 (1.966)	0.004	0.529
N-acetyl-1-methylhistidine*	Histidine Metabolism	Amino Acid	2.302 (0.980)	0.021	0.650
N-acetylcarnosine	Histidine Metabolism	Amino Acid	1.147 (0.476)	0.018	0.637
N-acetyltryptophan	Tryptophan Metabolism	Amino Acid	-23.570 (7.197)	0.001	0.424
N,N,N-trimethyl-alanylproline betaine (TMAP)	Urea cycle; Arginine and Proline Metabolism	Amino Acid	0.749 (0.333)	0.027	0.658
N1,N12-diacetylspermine	Polyamine Metabolism	Amino Acid	1.607 (0.782)	0.043	0.665
pro-hydroxy-pro	Urea cycle; Arginine and Proline Metabolism	Amino Acid	1.041 (0.377)	0.007	0.570
trans-urocanate	Histidine Metabolism	Amino Acid	-14.323 (3.917)	0.000 4	0.359
arabinose	Pentose Metabolism	Carbohydrate	1.293 (0.496)	0.011	0.570
2-methylcitrate	TCA Cycle	Energy	1.174 (0.448)	0.010	0.570
trans-aconitate	TCA Cycle	Energy	1.593 (0.661)	0.018	0.637

Metabolite	Sub-pathway	Super-pathway	Beta (SE)	*p*	*q*
glycerophosphoethanolamine	Phospholipid Metabolism	Lipid	1.450 (0.675)	0.034	0.662
glycerophosphoserine*	Phospholipid Metabolism	Lipid	0.729 (0.352)	0.041	0.665
2'-O-methyluridine	Pyrimidine Metabolism, Uracil containing	Nucleotide	1.359 (0.657)	0.041	0.665
3-(3-amino-3-carboxypropyl)uridine*	Pyrimidine Metabolism, Uracil containing	Nucleotide	0.640 (0.288)	0.029	0.658
3-ureidoisobutyrate	Pyrimidine Metabolism, Uracil containing	Nucleotide	4.110 (1.614)	0.013	0.570
3-ureidopropionate	Pyrimidine Metabolism, Uracil containing	Nucleotide	4.504 (1.890)	0.019	0.650
5,6-dihydrothymine	Pyrimidine Metabolism, Thymine containing	Nucleotide	1.570 (0.634)	0.015	0.625
5,6-dihydrouracil	Pyrimidine Metabolism, Uracil containing	Nucleotide	1.499 (0.497)	0.003	0.529
guanosine	Purine Metabolism, Guanine containing	Nucleotide	3.642 (1.434)	0.013	0.570
guanosine-3',5'-cyclic monophosphate (cGMP)	Purine Metabolism, Guanine containing	Nucleotide	1.492 (0.574)	0.011	0.570
hypoxanthine	Purine Metabolism, (Hypo)Xanthine/Inosine containing	Nucleotide	1.106 (0.431)	0.012	0.570
N4-acetylcytidine	Pyrimidine Metabolism, Cytidine containing	Nucleotide	0.599 (0.276)	0.033	0.662
uric acid ribonucleoside*	Purine Metabolism, (Hypo)Xanthine/Inosine containing	Nucleotide	0.949 (0.476)	0.049	0.665
xanthosine	Purine Metabolism, (Hypo)Xanthine/Inosine containing	Nucleotide	0.832 (0.353)	0.020	0.650
3-(4-hydroxyphenyl)propionate	Benzoate Metabolism	Xenobiotics	0.616 (0.277)	0.029	0.658
3-methoxycatechol sulfate (2)	Benzoate Metabolism	Xenobiotics	-5.785 (2.217)	0.011	0.570
erythritol	Food Component/Plant	Xenobiotics	0.731 (0.362)	0.046	0.665
salicylate	Drug - Topical Agents	Xenobiotics	2.738 (1.137)	0.018	0.637
X-12729	Uncharacterized	Uncharacterized	3.864 (1.736)	0.028	0.658
X-12822	Uncharacterized	Uncharacterized	0.780 (0.386)	0.046	0.665

Metabolite	Sub-pathway	Super-pathway	Beta (SE)	*p*	*q*
X-18887	Uncharacterized	Uncharacterized	1.226 (0.414)	0.004	0.529
X-23459	Uncharacterized	Uncharacterized	2.005 (0.960)	0.040	0.665
X-23593	Uncharacterized	Uncharacterized	0.808 (0.394)	0.043	0.665
X-23666	Uncharacterized	Uncharacterized	1.308 (0.515)	0.013	0.570
X-24411	Uncharacterized	Uncharacterized	-1.599 (0.587)	0.008	0.570
X-25983	Uncharacterized	Uncharacterized	-3.634 (1.606)	0.026	0.658

Beta beta-coefficient; *SE* standard error; *p* p-value; *q* q-value

Table S3.5. Summary statistics for metabolites associated with African genomic ancestry proportions (*p*<0.05) in infants of Hispanic/Latine maternal race/ethnicity, restricted to those also nominally significant in the primary analysis

Metabolite	Sub-pathway	Super-pathway	Beta (SE)	*p*	*q*
5-hydroxymethyl-2-furoic acid	Tyrosine Metabolism	Amino Acid	2.156 (0.847)	0.016	0.497
C-glycosyltryptophan	Tryptophan Metabolism	Amino Acid	1.327 (0.590)	0.032	0.616
dimethylguanidino valeric acid (DMGV)*	Urea cycle; Arginine and Proline Metabolism	Amino Acid	6.266 (1.921)	0.003	0.242
N-acetylcysteine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid	6.850 (1.871)	0.001	0.242
N-acetylglutamate	Glutamate Metabolism	Amino Acid	2.258 (0.850)	0.010	0.419
N-acetylmethionine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid	3.147 (1.461)	0.039	0.628
N-acetyltaurine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid	1.435 (0.564)	0.016	0.497
N-acetylvaline	Leucine, Isoleucine and Valine Metabolism	Amino Acid	1.443 (0.557)	0.014	0.497
arabinose	Pentose Metabolism	Carbohydrate	3.435 (1.007)	0.002	0.242
fructose	Fructose, Mannose and Galactose Metabolism	Carbohydrate	2.926 (1.221)	0.023	0.510
sucrose	Disaccharides and Oligosaccharides	Carbohydrate	35.627 (7.845)	$8x10^{-5}$	0.071

Metabolite	Sub-pathway	Super-pathway	Beta (SE)	*p*	*q*
glycerophosphoethanolamine	Phospholipid Metabolism	Lipid	3.189 (1.002)	0.003	0.252
glycerophosphoserine*	Phospholipid Metabolism	Lipid	1.692 (0.629)	0.011	0.447
dihydroorotate	Pyrimidine Metabolism, Orotate containing	Nucleotide	4.581 (1.451)	0.002	0.242
guanosine-3',5'-cyclic monophosphate (cGMP)	Purine Metabolism, Guanine containing	Nucleotide	4.927 (1.490)	0.002	0.242
N-carbamoylaspartate	Pyrimidine Metabolism, Orotate containing	Nucleotide	10.986 (5.214)	0.043	0.638
uric acid ribonucleoside*	Purine Metabolism, (Hypo)Xanthine/Inosine containing	Nucleotide	1.892 (0.772)	0.020	0.509
X-12729	Uncharacterized	Uncharacterized	6.335 (2.635)	0.022	0.510
X-18779	Uncharacterized	Uncharacterized	9.655 (3.933)	0.020	0.509
X-24228	Uncharacterized	Uncharacterized	2.143 (0.608)	0.001	0.242
X-24339	Uncharacterized	Uncharacterized	2.215 (0.757)	0.006	0.330
X-24403	Uncharacterized	Uncharacterized	6.217 (2.471)	0.017	0.497

Beta beta-coefficient; *SE* standard error; *p* p-value; *q* q-value

Table S3.6 Summary statistics for metabolites associated with proportions of African genomic ancestry at *p*<0.05 within infants of Black/African American maternal race/ethnicity

Available online as a separate file.

Table S3.7 Summary statistics for metabolites associated with proportions of African genomic ancestry at p<0.05 within infants of Hispanic/Latine maternal race/ethnicity

Available online as a separate file.

Chapter 4: Conclusions

This dissertation contributes to our understanding of metabolic dysregulation in extremely premature infants at high risk of developing bronchopulmonary dysplasia (BPD), and examines how ancestry-linked genetic variation shapes the urinary metabolome in this population. By integrating genomic and untargeted metabolomic data from two clinical cohorts of extremely premature infants, this work identified metabolomic patterns associated with BPD risk across the first month of life and investigated the contribution of genetic ancestry to metabolic heterogeneity. Collectively, these findings highlight the potential of multi-omic approaches to provide new insights into the biological processes underlying disease susceptibility and the factors contributing to inter-individual variability in the metabolome of this vulnerable population.

The findings presented in Chapter 2 demonstrate that metabolic dysregulation associated with BPD is dynamic, varying according to both postnatal age and feeding modality, with the most pronounced dysregulations observed during the fourth week of life. Although no individual metabolites remained significantly associated with BPD after correction for multiple testing, several biologically plausible candidate metabolites were identified, some of which have previously been suggested to be implicated in population-based, *in vivo*, and/or *in vitro* studies. Beyond individual metabolites, pathway-level analyses revealed distinct time- and feeding-modality-specific metabolic perturbations, including xanthine metabolism and enrichment of several acylcarnitine pathways, particularly among infants receiving parenteral nutrition during the fourth week of life. These results are consistent with observations reported in other studies of infant biofluids and support a broader role for altered energy metabolism in the development of BPD, highlighting this area as a candidate for future mechanistic and population-based studies. Of particular interest, a nominal, previously unreported pattern of vitamin dysregulation was also observed in infants who develop BPD. This highlights the need for further studies focused on nutritional supplementation, specifically among larger, more contemporary cohorts of extremely

premature infants, to characterize disease-specific metabolic dysregulation and to potentially identify candidate biomarkers for future clinical validation.

In Chapter 3, this dissertation transitioned from BPD-specific metabolic dysregulation to the broader question of how genetic variation captured by genomic ancestry proportions contributes to heterogeneity in the urinary metabolome of extremely premature infants. The work in this chapter demonstrates that African genomic ancestry contributes to variability in specific urinary metabolites throughout early life in infants born extremely premature. At the individual metabolite level, we identified urocanic acid isomers involved in histidine metabolism that were significantly associated with African genomic ancestry. At the pathway level, ancestry-associated metabolites at nominal levels were significantly enriched for functions in nucleotide metabolism, and depleted for functions in lipid metabolism. These findings serve as proof-of-concept that ancestry-linked genetic variation may contribute to variation in the metabolome of extremely premature infants, including metabolites of potential clinical relevance. Importantly, these genetic contributions are likely to differ across populations with varying genomic ancestry proportions, highlighting the importance of considering genomic ancestry in metabolomic studies to better understand sources of metabolic heterogeneity. Leveraging ancestry-linked genetic variation provides a powerful approach to identifying genetically regulated metabolites, particularly in smaller populations such as extremely premature infants, and may help identify metabolites for which the frequencies of regulatory alleles differ across continental populations. Future investigations are warranted, including admixture mapping in larger cohorts to identify the specific loci involved and to ultimately identify the specific ancestry-linked causal variation behind these associations.

The last objective of this dissertation aimed to determine if genetic ancestry-linked variation contributed to BPD-differential metabolites or metabolic pathways. While no formal statistical analyses were performed to evaluate this objective, we conducted a descriptive

evaluation of nominally associated metabolites within each study. Overall, we identified fourteen urinary metabolites that were nominally associated with BPD and that also varied by infant genetic ancestry (Table 4.1), including five metabolites involved in amino acid metabolism. However, the overlap between the two analyses was not statistically significant (Fisher's exact test, $p = 0.15$), suggesting no evidence of enrichment beyond what would be expected by chance. Although these qualitative observations do not provide evidence of enrichment beyond what would be expected by chance, they highlight the importance of considering both the causal and confounding roles of genetic and environmental contributions to the urinary metabolome when studying disease susceptibility in diverse populations of extremely premature infants. Future analyses using multi-omic approaches offer a promising path to unravel the complexity of BPD susceptibility among extremely premature infants, a vulnerable pediatric population that is commonly overlooked in scientific studies.

Table 4.1. Intersection of metabolite associations at $p<0.05$ with bronchopulmonary dysplasia in either time point and African genomic ancestry proportions

		Chapter 2						Chapter 3		
		Time Point 1			Time Point 2			African Genomic Ancestry		
Metabolite	**Super Pathway**	**Beta**	**SE**	***p***	**Beta**	**SE**	***p***	**Beta**	**SE**	***p***
5-(galactosylhydroxy)-lysine	Amino Acid	-0.077	0.227	0.735	-0.545	0.185	0.003	0.759	0.284	0.008
5-hydroxyindoleacetate	Amino Acid	-0.001	0.192	0.996	-0.345	0.168	0.040	0.840	0.317	0.009
hydroxy-N6,N6,N6-trimethyllysine*	Amino Acid	-0.069	0.229	0.763	-0.578	0.223	0.009	0.784	0.266	0.003
N-acetyl-1-methylhistidine*	Amino Acid	-0.134	0.100	0.181	-0.357	0.151	0.018	1.828	0.619	0.003
pro-hydroxy-pro	Amino Acid	0.100	0.198	0.612	-0.351	0.157	0.026	0.695	0.297	0.020
N-acetylglucosaminylasparagine	Carbohydrate	0.200	0.281	0.477	-0.436	0.211	0.039	0.561	0.242	0.021
sucrose	Carbohydrate	-0.076	0.029	0.010	-0.036	0.028	0.204	6.220	2.811	0.028
N,N-dimethyl-pro-pro	Peptide	-0.124	0.316	0.695	-0.585	0.275	0.033	0.452	0.205	0.028
X-12708	Uncharacterized	-0.178	0.274	0.515	-0.475	0.220	0.031	0.533	0.250	0.034
X-12822	Uncharacterized	-0.114	0.238	0.631	-0.618	0.229	0.007	0.626	0.284	0.028
X-18838	Uncharacterized	-0.134	0.078	0.085	-0.251	0.087	0.004	2.032	0.758	0.008
X-22152	Uncharacterized	-0.192	0.170	0.258	-0.320	0.146	0.029	0.797	0.367	0.031
X-25457	Uncharacterized	-0.134	0.082	0.103	-0.362	0.104	0.001	1.973	0.690	0.005
erythritol	Xenobiotics	-0.060	0.246	0.807	-0.517	0.211	0.014	0.717	0.261	0.006

Notes: This table presents descriptive summary statistics for metabolites with $p<0.05$ in both sets of analyses, rather than results of a new analysis. Values under Chapter 2 correspond to the analyses using a logistic regression model for the association between metabolite levels and bronchopulmonary dysplasia at two time points: Time Point 1 (5-15 postnatal days) and Time Point 2 (23-32 postnatal days). Values under Chapter 3 correspond to the linear mixed models used to determine the association between African genomic ancestry proportions and metabolite levels.

Abbreviations: *Beta* beta coefficient; SE standard error; *p* *p*-value

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

Signed by:
Maybeth L. Malave-Mendez
02DC850C58004F7...
Author Signature

08/17/2026
Date