

## ARTICLE OPEN



# Distinct metabolic profiles associated with autism spectrum disorder versus cancer in individuals with germline *PTEN* mutations

Lamis Yehia<sup>1,8</sup>, Ying Ni<sup>2,8</sup>, Tammy Sadler<sup>1</sup>, Thomas W. Frazier<sup>1,3,4</sup> and Charis Eng<sup>1,5,6,7</sup>✉

*PTEN* hamartoma tumor syndrome (PHTS), caused by germline *PTEN* mutations, has been associated with organ-specific cancers and autism spectrum disorder (ASD) and/or developmental delay (DD). Predicting precise clinical phenotypes in any one PHTS individual remains impossible. We conducted an untargeted metabolomics study on an age- and sex-matched series of PHTS individuals with ASD/DD, cancer, or both phenotypes. Using agnostic metabolomic-analyses from patient-derived lymphoblastoid cells and their spent media, we found 52 differentially abundant individual metabolites, 69 cell/media metabolite ratios, and 327 pair-wise metabolite (shared metabolic phenotype) ratios clearly distinguishing PHTS individuals based on phenotype. Network analysis based on significant metabolites pointed to hubs converging on *PTEN*-related insulin, MAPK, AMPK, and mTOR signaling cascades. Internal cross-validation of significant metabolites showed optimal overall accuracy in distinguishing PHTS individuals with ASD/DD versus those with cancer. Such metabolomic markers may enable more accurate risk predictions and prevention in individual PHTS patients at highest risk.

npj Genomic Medicine (2022)7:16; <https://doi.org/10.1038/s41525-022-00289-x>

## INTRODUCTION

Hereditary cancer predisposition syndromes and neurodevelopmental disorders account for a large subset of individuals in the medical genetics clinic<sup>1–3</sup>. While advances in genomic medicine have enabled the identification of the underlying etiologies for many of these disorders, genotype–phenotype associations are not absolute; it remains challenging to predict the natural history of any one hereditary disorder at the individual patient level versus the population/cohort level<sup>4</sup>. One well-studied model is *PTEN* hamartoma tumor syndrome (PHTS, MIM 158350), a spectrum of cancer- and neurodevelopmental disorders-related phenotypes caused by germline mutations in the tumor suppressor phosphatase and tensin homolog gene (*PTEN*, MIM 601728)<sup>5,6</sup>. While *PTEN* germline mutations were originally identified in a relatively rare subset of disorders predisposing to breast, thyroid, and other cancers<sup>7</sup>, subsequent studies have shown that *PTEN* germline mutation is amongst the most common causes of autism spectrum disorder (ASD)<sup>8,9</sup>. This *PTEN*-related phenotypic dichotomy poses a challenge for more timely and precise medical management of individuals with germline *PTEN* mutations<sup>6,10</sup>. Deciphering this dichotomy may also have value in identifying the etiology of a subset of ASD/DD.

The extensive phenotypic heterogeneity in PHTS supported the hypothesis that genetic or genomic modifying factors exist. First, earlier studies showed that germline variants in genes encoding three of the four subunits of succinate dehydrogenase or mitochondrial complex II (*SDHB*, *SDHC*, and *SDHD*, collectively referred to as *SDHx*) modify breast cancer risk and thyroid cancer histology in individuals with germline *PTEN* mutations<sup>11,12</sup>. Second, copy number variations were found to be associated

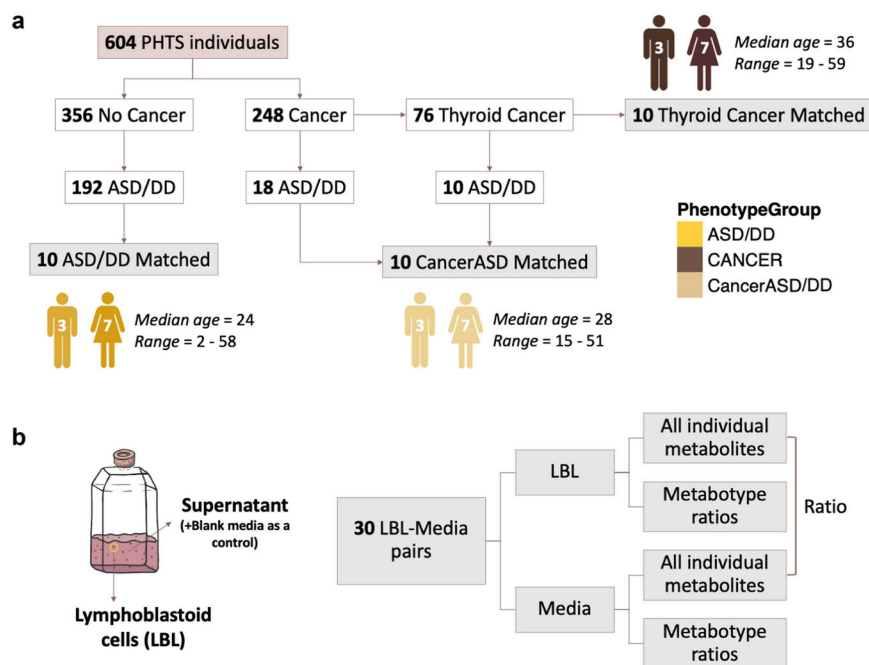
with the ASD and/or developmental delay (DD) phenotype versus cancer in patients with germline *PTEN* mutations<sup>13</sup>. These two studies provided proof-of-principle that genomic modifiers may play a role in modulating phenotypic outcomes in PHTS.

Metabolomics, the comprehensive study of small molecules, known as metabolites, in biological systems, has emerged as a promising analytical profiling method for biomarker discovery<sup>14</sup>. Multiple studies have elucidated the association of multiple metabolites and/or metabolic pathways in pathobiological processes, including sporadic cancer and nonsyndromic neurodevelopmental disorders<sup>15–22</sup>. However, the variability among metabolites and metabolic pathways implicated in the same phenotype only reflects the complexity and heterogeneity of such disorders. Recently, metabolotyping, a subtyping approach based on shared metabolic phenotypes identified from a set of metabolic biomarkers has shown promise in screening for autism risk in children<sup>17,23</sup>.

Relevant to PHTS and the known role of mitochondrial energetics in this syndrome, we had previously conducted a pilot study using a targeted approach that focused on metabolites within the tricarboxylic acid cycle<sup>16</sup>. This study provided proof-of-principle regarding the role of metabolites as predictive markers of phenotypic outcomes in PHTS. However, because these studies have been limited to a small targeted set of metabolites, the associations we identified thus far represent an incomplete snapshot of metabolites that may influence ASD/DD and/or cancer outcomes in individuals with PHTS.

Integrating metabolomic profiles on top of the genetic predisposition for phenotype prediction is of great interest, especially for inherited disorders like PHTS with seemingly

<sup>1</sup>Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>2</sup>Center for Immunotherapy and Precision Immuno-Oncology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>3</sup>Autism Speaks, Cleveland, OH, USA. <sup>4</sup>Department of Psychology, John Carroll University, University Heights, OH, USA. <sup>5</sup>Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. <sup>6</sup>Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. <sup>7</sup>Germline High Risk Cancer Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. <sup>8</sup>These authors contributed equally: Lamis Yehia, Ying Ni. ✉email: [engc@ccf.org](mailto:engc@ccf.org)



**Fig. 1** Characteristics of study participants and study design. **a** We selected a series of 30 sex- and age-matched PHTS individuals for untargeted metabolomics measurements. **b** Metabolite sample measurements were obtained from lymphoblastoid cell lines (LBL), paired surrounding growth media for each sample, and blank growth media serving as a baseline negative control. PHTS PTEN hamartoma tumor syndrome, ASD autism spectrum disorder, DD developmental delay.

disparate cancer and neurodevelopmental phenotypes. As such, we sought to address the hypothesis that specific metabolites and/or metabolic networks in patients carrying germline *PTEN* mutations are associated with the development of specific clinical phenotypes, here, cancer versus ASD/DD. Thus, we performed a comprehensive untargeted metabolomics approach in a matched series of PHTS individuals.

## RESULTS

### Research participants and study design

Six hundred and four individuals diagnosed with PHTS were accrued based on genotypic and phenotypic characteristics and under the approved research protocol 8458-PTEN at the Cleveland Clinic. Of those, 248 (41%) had at least one cancer diagnosis. In this study, we limited the cancer diagnosis to only that of the thyroid since this cancer is the earliest onset cancer type in PHTS affecting both males and females (youngest age at diagnosis starting at 7 years)<sup>24,25</sup>, and importantly, to minimize variability in a deliberately smaller sample size. The earlier age at onset for differentiated thyroid cancer also enables appropriate matching of samples with other young PHTS individuals who have ASD and/or DD. The final matched series consisted of 30 eligible individuals with PHTS (Fig. 1a and Supplementary Table 1). Both females (70%) and males (30%) were represented, and the median age at consent was 28 years (range 2–59). Research participants were divided into three age- and sex-matched groups. The first group consisted of 10 PHTS individuals diagnosed with ASD and/or DD without cancer identified to date, with a median age at consent of 24 years (range 2–58). The second group consisted of 10 PHTS individuals diagnosed with thyroid cancer (median age at consent = 36 years; range = 19–59), with a median age at cancer diagnosis of 25 years (range 17–41). The third group consisted of 10 PHTS individuals (median age at consent = 28 years; range = 15–51) who in addition to ASD/DD, had a cancer diagnosis, with a median age at onset of 18 years (range 7–51). Of note, 7 of the 10 cancer diagnoses consisted of thyroid cancer (Supplementary Table 1).

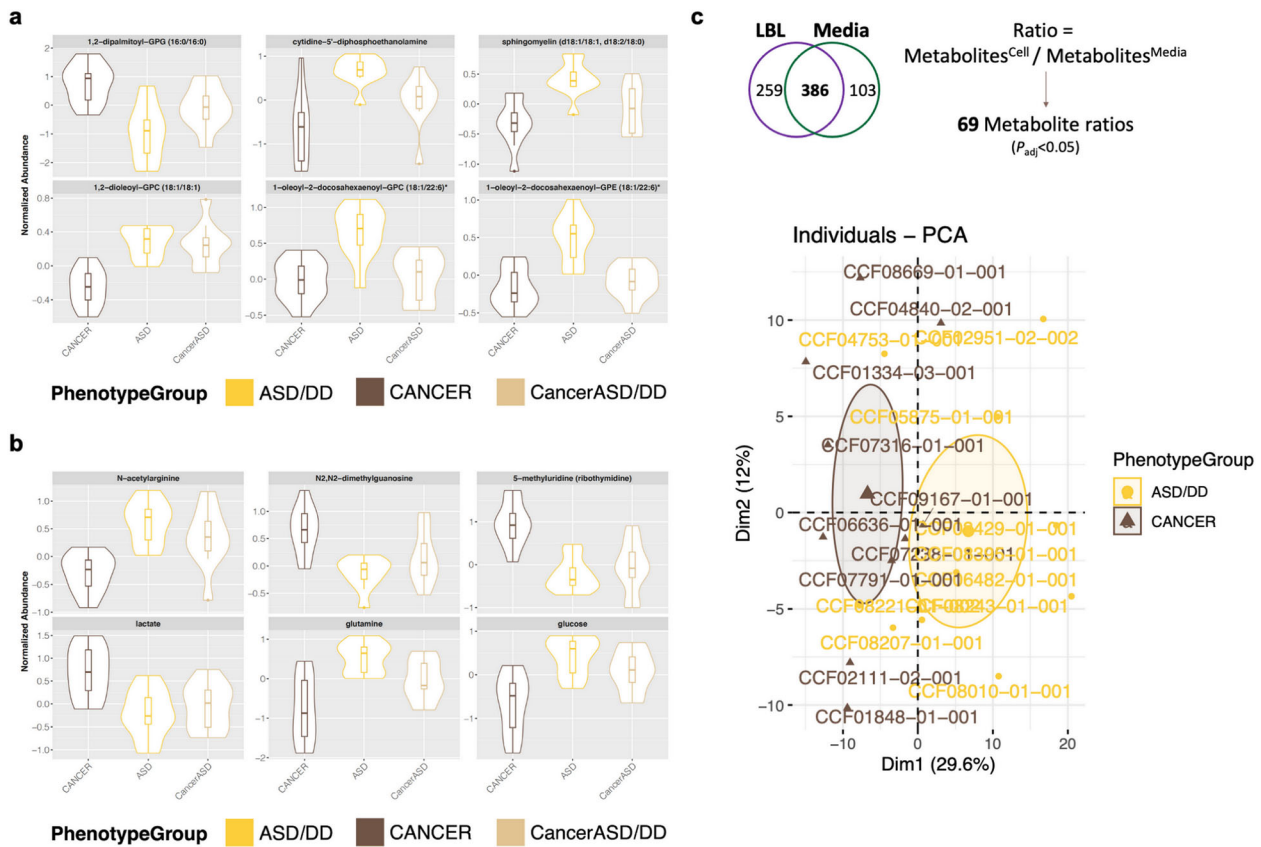
The three phenotype groups were matched with respect to biological sex ( $P = 1$ , Fisher's exact test) and ages at consent ( $P = 0.38$ , effect size =  $-0.002$ , Kruskal–Wallis rank sum test).

We generated immortalized lymphoblastoid cell lines (LBLs) from each research participant following standard procedures and then cultured them in vitro for metabolic profiling. Metabolite sample measurements were obtained from LBLs, paired surrounding growth media for each sample, and blank growth media serving as a baseline negative control. We performed analyses examining individual metabolites in each matrix (cells and media), ratio of metabolites shared between the two matrices, and metabotyping analysis focusing on biologically-relevant metabolites (Fig. 1b).

### Identification of single differentially abundant metabolites

Quantified metabolites from LBLs and media consist of amino acids, carbohydrates, cofactors and vitamins, those related to energy metabolism, lipids, nucleotides, peptides, xenobiotics, and partially characterized molecules. Collectively, we detected 645 metabolites from LBLs and 489 metabolites from their matched spent media. As expected, the constitution of measured metabolites differed between the two biological matrices, with enrichment of lipids in the cell compartment compared to the media (Supplementary Fig. 1).

We identified multiple differentially abundant metabolites (adjusted  $P < 0.05$ ) amongst samples derived from PHTS individuals with ASD/DD, cancer, and both phenotypes (Supplementary Fig. 2). Using the cellular compartment, we detected multiple differentially abundant metabolites amongst cells derived from PHTS individuals belonging to the three phenotype groups (Fig. 2a and Supplementary Fig. 2). These metabolites generally belong to lipid (89%) and amino acid (11%) classes of molecules (Supplementary Data 1). Similarly, we detected multiple differentially abundant metabolites in the surrounding media of cells derived from PHTS individuals (Fig. 2b and Supplementary Fig. 2). Here, metabolites predominantly belong to amino acids (54%) followed by nucleotides (29%), carbohydrates (8%), cofactors and



**Fig. 2 Identification of differentially abundant individual metabolites.** **a** Top six differentially abundant metabolites (adjusted  $P < 0.05$ ) amongst cells derived from PHTS individuals belonging to the three phenotype groups. **b** Top six differentially abundant metabolites (adjusted  $P < 0.05$ ) in the surrounding media of cells derived from PHTS individuals. For the violin plots, the center lines represent the median. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest value no further than  $1.5 \times$  inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most  $1.5 \times$  IQR of the hinge. **c** There are 386 metabolites shared in both the cellular and media matrices. Comparisons of the ratios of these metabolites resulted in 69 differentially abundant metabolite ratios (adjusted  $P < 0.05$ ) that successfully separated PHTS individuals according to phenotype group.

vitamins (4%), and xenobiotics (4%) (Supplementary Data 1). The principal component analysis (PCA) score plot of metabolic profiles from the media compartment shows clear separation from the blank growth media, serving as a negative control (Supplementary Fig. 3).

Based on our observations interrogating each of the biological matrices (cells and media) on its own (Supplementary Data 2), we then evaluated the ratio of metabolites in cells over spent media, since this method of analysis may uncover biologically-relevant metabolomic alterations that otherwise would not be able to be captured with single metabolite analyses<sup>23,26</sup>. This analysis only applied to metabolites detected in both matrices (Fig. 1b), which led to 386 shared metabolites. There are 69 metabolite cells/media ratios that are significantly different between PHTS individuals with ASD/DD and cancer phenotypes. This analysis revealed a clear separation of PHTS individuals according to phenotype group (Fig. 2c).

### Pathway enrichment analysis

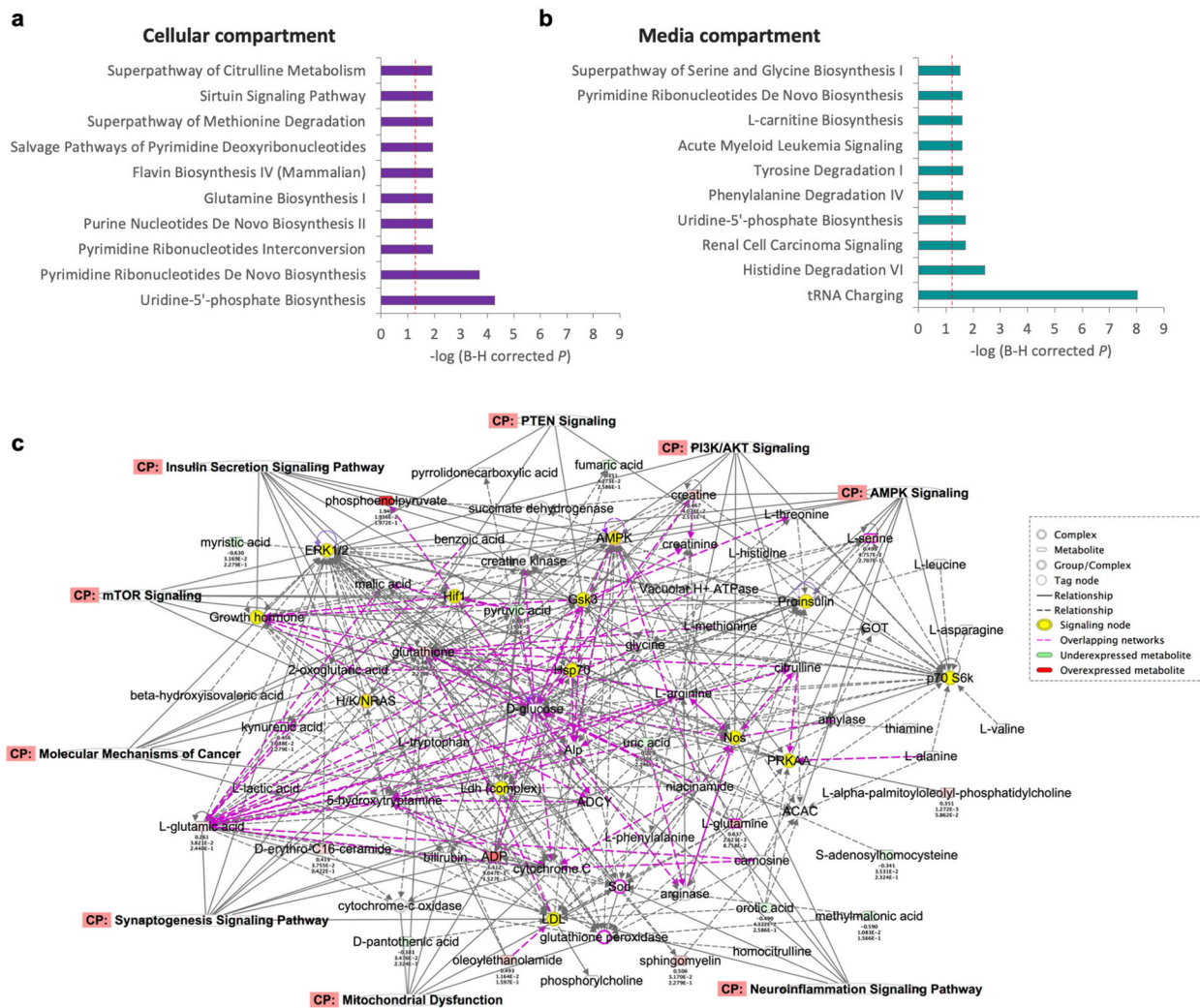
To elucidate the biological effects of the observed metabolomic alterations as relevant to cancer and ASD/DD, we performed pathway enrichment analysis. Using the cellular compartment, the top five most enriched canonical pathways comparing ASD/DD versus cancer included glutamine biosynthesis ( $P = 1.2 \times 10^{-2}$ ), and broadly implicated nucleotide metabolism ( $P \leq 1.2 \times 10^{-2}$ ) (Fig. 3a). The top molecular network focusing on diseases and

functions implicated cellular compromise, lipid metabolism, and small molecule biochemistry ( $P = 10^{-44}$ ). Using the surrounding media, we identify a predominantly amino acid enriched canonical pathway signature, with tRNA charging ( $P = 9.8 \times 10^{-9}$ ) as the top significant pathway (Fig. 3b). The top molecular network implicated amino acid metabolism, molecular transport, and small molecule biochemistry ( $P = 10^{-60}$ ). Finally, the cells/media ratios analysis resulted in only three significantly enriched canonical pathways, including tRNA charging ( $P = 4.3 \times 10^{-8}$ ), glycine betaine degradation ( $P = 3 \times 10^{-3}$ ), and glycine biosynthesis ( $P = 7.8 \times 10^{-3}$ ) (Supplementary Data 3). Similar to the media analysis, the top molecular network included amino acid metabolism, molecular transport, and small molecule biochemistry ( $P = 10^{-46}$ ). Intriguingly, the molecular networks from all three analyses converged on the PTEN-relevant insulin signaling pathway, as well as ERK1/2, AMPK, and mTOR signaling cascades ( $P \leq 10^{-44}$ ) (Fig. 3c).

### Metabotype clusters can distinguish PHTS patients with cancer versus those with autism

We also implemented a metabotyping approach, subtyping based on shared metabolic phenotypes that has shown promise in risk stratification, including for ASD<sup>17</sup>. Accordingly, guided by our pilot studies and metabolic pathways known to be disrupted in ASD generally<sup>16,17,27,28</sup>, we assessed 43 metabolites associated with amino acid metabolism and mitochondrial energetics in PHTS





**Fig. 3** Pathway enrichment analysis and convergence on PTEN-related networks. **a** Top ten most enriched canonical pathways associated with differentially abundant metabolites detected from the cellular compartment. **b** Top ten most enriched canonical pathways associated with differentially abundant metabolites detected from the media compartment. Dashed red lines in **(a, b)** indicate the threshold for a significant corrected  $P$  value  $< 0.05$ . **c** Molecular networks integrating all differentially abundant metabolites converge on the PTEN-relevant insulin signaling pathway, as well as ERK1/2, AMPK, and mTOR signaling cascades ( $P \leq 10^{-44}$ ).

individuals belonging to the three phenotype groups (Supplementary Table 2). The pair-wise correlation analysis resulted in a total of 703 ratio tests from 38 metabolites detected in the cellular matrix, and 741 ratio tests from 39 metabolites detected in the media matrix. Using the media matrix, we identified 175 differentially underexpressed and 152 overexpressed ratio tests in PHTS individuals with ASD/DD relative to PHTS individuals with cancer (Fig. 4a).

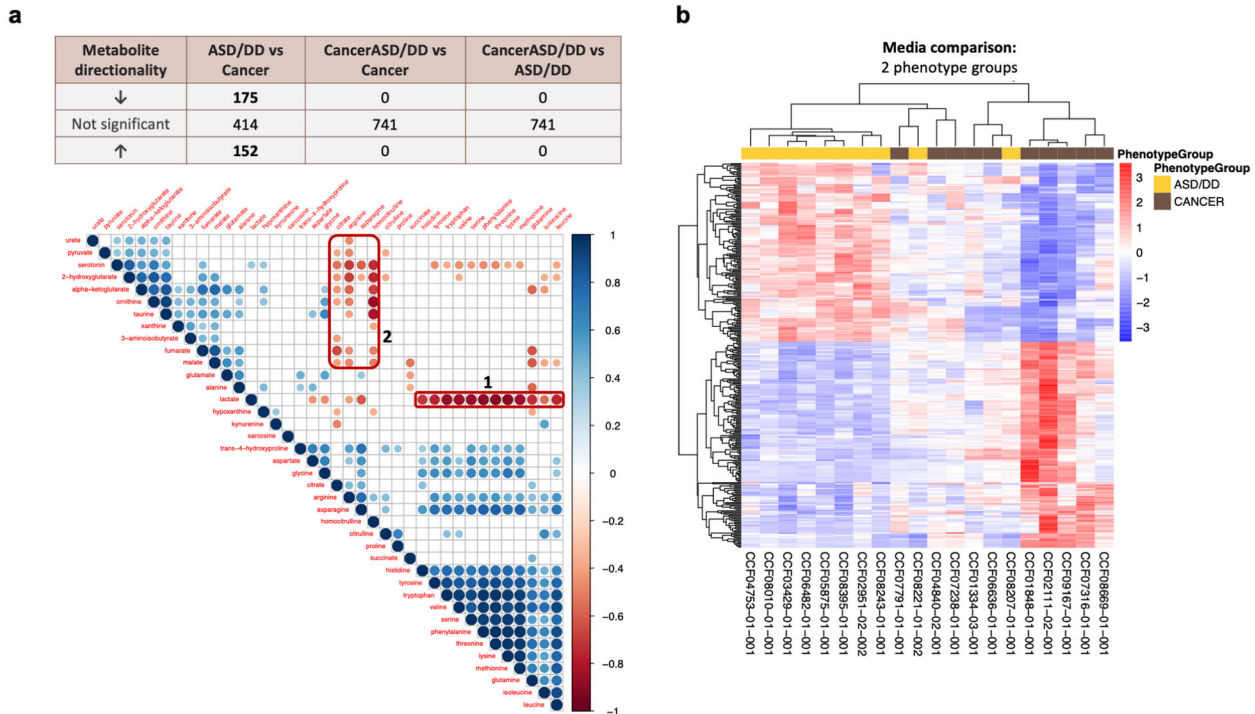
As the negative correlation represents the potential highest pair-wise ratio signal, we observed the most prominent negative correlation blocks for metabolites detected from the media compartment, including lactate versus the amino acid group (block 1) and between arginine, homocitrulline versus serotonin, hydroxyglutarate, ornithine, taurine (block 2) (Fig. 4a). Differential abundance analysis adjusted for age and sex did not identify any significant differentially abundant metabolite clusters in the cellular matrix. As expected, no clear negative correlation block was observed from metabolites detected from cells (Supplementary Fig. 4).

Unsupervised hierarchical clustering shows that significant differentially abundant metabolotypes from media distinctly separate PHTS individuals with cancer from those who have ASD/DD,

except for two patients (Fig. 4b). Of note, we applied correlation distance measurement and complete clustering methods to delineate group similarities among all individuals included in the heatmap, rather than for classification purposes. The first patient, CCF08221-01-002, is a 2-year-old female with macrocephaly, ASD, global DD, and café-au-lait spots. CCF08207-01-001 is a 46-year-old female with macrocephaly, DD, oral mucosa papilloma, atypical ductal breast hyperplasia, breast fibrocystic disease, intraductal papilloma of breast, noninfiltrating intraductal and lobular carcinoma of the breast, goiter, skin tag, and ovarian cysts.

### Performance of different metabolite predictors

To evaluate the predictive value of metabolite differences we identified to discriminate among phenotypes, we performed internal cross-validation of significant metabolites using the leave one out cross-validation (LOOCV) approach on our relatively small-sized but well-matched samples. This cross-validation indicated optimal sensitivity, specificity, and overall accuracy in distinguishing PHTS individuals with ASD/DD versus those who have cancer using agnostic metabolomic measurements. We compared the classification performance using significant differentially abundant metabolites identified from LBL, media, LBL/media ratio, as well as



**Fig. 4 Metabolite clusters can distinguish individuals with cancer versus those with ASD/DD.** **a** Differential abundance analysis adjusted for age and sex identified differentially underexpressed and overexpressed metabolite ratio tests in media from samples of PHTS individuals with ASD/DD relative to PHTS individuals with cancer. Correlation plot showing prominent negative correlation blocks (red rectangles). Circles shown are for results significant at  $P < 0.05$ , with increasing diameter/color corresponding with increasing correlation (circles omitted otherwise). **b** Using the media compartment, unsupervised hierarchical clustering shows that significant differentially abundant metabolotypes distinctly separate PHTS individuals with cancer from those who have ASD/DD, with the exception of two patients. ASD autism spectrum disorder, DD developmental delay.

**Table 1.** Performance metrics.

Matrix	ASD/DD vs. cancer ("positive" class: cancer)	Overall accuracy	Sensitivity	Specificity
Media	AllSigMetabolites ( $N = 23$ )	0.77	0.73	0.80
	RemoveLinearCorrelated ( $N = 20$ )	0.77	0.73	0.80
	PCA ( $N = 7$ )	0.67	0.60	0.73
Cell	AllSigMetabolites ( $N = 8$ )	0.80	0.80	0.80
	RemoveLinearCorrelated <sup>a</sup> ( $N = 8$ )	0.80	0.80	0.80
	PCA ( $N = 6$ )	<b>0.95</b>	<b>0.90</b>	<b>1.00</b>
Individual metabolite ratio	AllSigMetabolites ( $N = 69$ )	0.80	0.90	0.70
	RemoveLinearCorrelated ( $N = 20$ )	<b>0.93</b>	<b>1.00</b>	<b>0.87</b>
	PCA ( $N = 11$ )	0.87	0.90	0.83
Metabolite ratio	AllSigMetabolites ( $N = 327$ )	0.62	0.63	0.60
	RemoveLinearCorrelated ( $N = 20$ )	0.80	0.80	0.80
	PCA ( $N = 9$ )	0.77	0.80	0.73

AllSigMetabolites all significantly abundant metabolites, PCA principal component analysis, ASD autism spectrum disorder, DD developmental delay.

<sup>a</sup>No linear correlated component to be removed. Bolded numbers refer to the comparisons with the highest overall accuracy.

metabolite ratio, with or without dimension reduction by PCA analysis or removing linearly correlated metabolites. In general, 2-group (Cancer versus ASD) classification performed better than 3-group (Cancer, ASD, and CancerASD) (Supplementary Table 3). In particular, 6 PCA components derived from LBL differentially abundant metabolites showed the highest overall accuracy (0.95, 95% CI: 0.86–0.99,  $P = 3.1 \times 10^{-14}$  compared to No Information Rate [NIR]) distinguishing samples in the ASD/DD group from the cancer group, with a sensitivity of 0.90 and specificity of 1.

Metabolite ratios between LBL/media show the second best overall accuracy (0.93, 95% CI: 0.84–0.98,  $P = 4.5 \times 10^{-13}$  compared to NIR) after removing linear correlated ratios, with a sensitivity of 1, and specificity of 0.87 (Table 1).

To identify group similarities in metabolomic profiles amongst PHTS individuals with different phenotypes, unsupervised hierarchical clustering of significantly abundant metabolites derived from the LBL component showed that these metabolites distinctly separate PHTS individuals with cancer, with the exception of two





biologically-relevant alterations with predictive value at the individual level.

The Children's Autism Metabolome Project (CAMP, clinical trial NCT02548442) recently found and replicated plasma metabolotypes in a series of 499 children with *idiopathic* ASD and 209 typically developing children<sup>17</sup>. Metabotyping is a subtyping approach based on shared metabolic phenotypes identified from a set of metabolic biomarkers<sup>17,23</sup>. CAMP prioritized 39 ASD-focused metabolites associated with amino acid and energy metabolism to identify 34 metabolotypes that could detect >50% of autistic participants, which may have clinical value. In this study, we hypothesized that the established ASD metabolotypes<sup>17,23</sup> may have a predictive value in distinguishing PHTS individuals with ASD/DD versus those without ASD/DD, in the research setting. While our data showed acceptable separation of PHTS individuals by ASD/DD versus cancer phenotypes based on metabolotype clusters, the overall sensitivity, specificity, and accuracy were lower than the power of individual metabolites from the cellular matrix and the ratio analysis (Table 1). CAMP focused on children 18–48 months of age and quantified metabolotypes from plasma<sup>17</sup>. Because we utilized LBLs and their media for metabolite measurements and included PHTS individuals regardless of age, it is plausible that the latter may have metabolotype differences involving different metabolite classes than identified by CAMP. Importantly, in our comparisons, all individuals harbored germline *PTEN* mutations, which might obscure signals expected when comparing these PHTS individuals to individuals who are *PTEN* wildtype<sup>16</sup>. These speculations warrant further investigation with more burgeoning data regarding the pathobiology of *PTEN*-related ASD/DD. Importantly, external replication of our internal cross-validation approach will be important for ensuring that small variations in sampling are not significantly impacting model performance in prospective cases.

One intriguing observation is the presence of “misclassified” individuals upon unsupervised hierarchical mapping based on significantly abundant metabolite clusters (Figs. 4b, 5). We speculate that these PHTS-ASD/DD individuals clustered with the PHTS-Cancer group may be at the highest risk of developing cancer during their lifetimes, and longitudinal studies will help test this long-term but important hypothesis. Amongst the three PHTS-ASD/DD individuals clustered with the PHTS-Cancer group using significant abundant metabolites from all analyses (cells, media, and metabolotypes), two had a family history of cancer, whereas this information was unknown for the third. However, family history of cancer cannot explain the distinct clustering relative to the other PHTS-ASD/DD individuals forming a uniform cluster. Indeed, of the remaining seven PHTS-ASD/DD individuals, 4 out of 5 with family history information had a positive family history of cancer. Therefore, our findings support our previous observations that family history of cancer alone is not predictive of personal cancer history at the individual level<sup>5</sup>. The alternative, but not mutually exclusive, explanation may be that these cluster “misclassified” ASD/DD individuals are destined to develop malignancies in adulthood. Indeed, one of these PHTS-ASD/DD individuals (CCF08207-01-001) was found to have ductal and lobular carcinoma in situ of the breast, a pre-invasive breast cancer. Conversely, only one PHTS research participant with cancer clustered with the ASD/DD group. This is a 24-year-old female with thyroid cancer, goiter, Hashimoto's disease, and hemangiomas. To our knowledge, she has not had a formal evaluation for ASD. Hence, we posit that such an approach can help screen for individuals at the highest risk of either phenotype. Unexpectedly, PHTS individuals with both cancer and ASD/DD clustered indiscriminately between the PHTS-Cancer and PHTS-ASD/DD groups. This suggests that this rare subset of PHTS individuals may represent a unique biological subgroup, at least as related to metabolomic profiles.

Overall, this study provides further robust evidence for the roles of relevant metabolites in PHTS-ASD/DD versus PHTS-Cancer etiologies, with the added advantage of having a more homogeneous (monogenic) and well-matched sample series through our study design. Despite research showing the extensive overlap in risk genes (one being *PTEN*) and biological pathways for ASD and for cancer<sup>32,33</sup>, studying such homogeneous patient series will help identify the intricate differences that define the dichotomous phenotype context. The ability to utilize such metabolomic markers will provide a clinical translational framework for stratifying *individual* PHTS patients based on more accurate ASD/DD versus cancer risk predictions, enabling precise risk assessment and early intervention in those at the highest risk.

## METHODS

### Research participants and clinical data

A total of 604 individuals clinically diagnosed with *PTEN* hamartoma tumor syndrome were prospectively accrued in accordance with research protocol 8458-*PTEN*, approved by the Cleveland Clinic Institutional Review Board. To address our research question, we prioritized an age- and sex-matched series of 30 individuals with ASD and/or DD without a personal history of invasive cancer(s) (excluding Stage 0 noninfiltrating intraductal and/or lobular carcinoma of the breast), differentiated thyroid cancer, and those with ASD/DD in addition to a cancer diagnosis (majority having thyroid cancer). For each consented research participant, we reviewed medical records, including clinical genetic testing reports, pedigrees, clinical notes associated with cancer genetics and/or genetic-counseling visits, and ASD Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria, where applicable. Written informed consents were obtained from all research participants.

### *PTEN* mutation and deletion analysis

Germline genomic DNA samples from peripheral blood leukocytes were extracted by the Genomic Medicine Biorepository (GMB) of the Cleveland Clinic Genomic Medicine Institute (Cleveland, OH, USA) using standard methods (<https://www.lerner.ccf.org/gmi/gmb/>). *PTEN* mutation and deletion analysis were performed as previously reported<sup>34</sup>. Mutation analysis was performed with a combination of denaturing gradient gel electrophoresis (DGGE), high-resolution melting curve analysis, and direct Sanger sequencing (ABI 3730xl; Applied Biosystems, Life Technologies) (Supplementary Table 4). Deletion analysis was performed using the multiplex ligation-dependent probe amplification kit (P158; MRC-Holland) according to manufacturer protocol. All patients underwent polymerase chain reaction-based Sanger sequencing of the *PTEN* promoter region. For *PTEN* germline variant positive individuals, pathogenicity predictions are reported according to orthogonal testing in a CLIA-certified facility, ClinVar database classifications, and/or the ClinGen gene-specific criteria for *PTEN* variant curation<sup>35</sup>. Carriers of *PTEN* promoter variants were considered as mutation positive only if the underlying variants have been associated with PHTS or known to affect *PTEN* function<sup>25,36–38</sup>.

### Cell lines and culture conditions

Immortalized LBLs from peripheral blood samples of individuals with PHTS were generated by the GMB (Cleveland, OH, USA) following standard procedures (<https://www.lerner.ccf.org/gmi/gmb/>). Cells were subsequently grown in RPMI-1640 supplemented with 20% fetal bovine serum and 1% penicillin/streptomycin and maintained at 37 °C and 5% CO<sub>2</sub> culture conditions. All cell lines remained anonymized and devoid of any identifiers (only number coded) during the duration of the experiments.

### Sample processing and metabolite measurement template preparation

We seeded the non-adherent LBLs at a density of 10 million cells per T75 flask. At the time of seeding, we transferred 1.5 ml of growth media into three independent aliquots (500 µl each). These media aliquots represent the ‘blank’ metabolite measurements to account for compounds already present in the cell culture media. Cells were allowed to grow overnight and subsequently collected into 50 ml conical tubes. We spun down the cell suspension at 1000 RPM for 5 min in a cooled centrifuge (4 °C). We transferred 1 ml of supernatant from each cell line and saved in fresh and

cooled 1.5 ml tubes. These supernatants represent the “secretome” portion of our untargeted metabolomics analysis. We added 1 ml of ice-cold PBS to wash the cell pellet. The cell suspension was then transferred to 1.5 ml tubes, spun down at 1000 RPM for 5 min at 4 °C to remove the PBS wash supernatant. This resulted in ~100 µl of packed cell pellet per sample. All samples (media and cell pellets) were flash frozen using liquid nitrogen and stored at –80 °C.

### Measurement of global untargeted metabolites

Untargeted metabolomics measurements were conducted at Metabolon (Morrisville, NC, USA) using ultrahigh performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS)<sup>39</sup>. Following receipt by the GMB, samples were inventoried and immediately stored at –80 °C. Following standard procedures to recover metabolites, the resulting extract from each sample was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. This strategy ensured maximal recovery and coverage of metabolites. All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high-resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. Sample extracts were dried, then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained standards at fixed concentrations to ensure injection and chromatographic consistency. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon’s library is based on authenticated standards that contains the retention time/index (RI), mass to charge ratio ( $m/z$ ), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Additional mass spectral entries have been created for structurally unnamed biochemicals, identified by their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis. For the purposes of this study, unknown compounds were excluded from all downstream analyses.

### Data normalization and statistical analyses

For each metabolite, the raw values in the experimental samples are divided by the median of those samples in each instrument batch, giving each batch and thus the metabolite a median of one. The minimum value across all batches in the median scaled data is imputed for the missing values. Metabolites measured from cell pellets are first batch normalized and then divided by the protein concentration, before re-scaling to have median = 1 (divide the new values by the overall median for each metabolite). Afterward, imputation is performed. The batch normalized and imputed data were transformed using the natural log. Final log-transformed and center-scaled metabolite values have a mean of 0 and a standard deviation of 1.

Statistical analyses were conducted with RStudio version 1.4.1717 (<https://www.rstudio.com>). The linear model was applied to assess differential metabolite expression in the context of a multifactor designed experiment with Limma R package version 3.48.0<sup>40,41</sup>, using age at consent and biological sex as covariates. Furthermore, statistically significant metabolic features were extracted using the criteria of multiple testing corrected  $p$  value <0.05 and visualized using volcano plots and heatmaps. For pair-wise comparisons, we used two-sided Student’s  $t$  tests and/or Wilcoxon’s rank sum tests, as indicated. Statistically significant metabolites were then used for networking analysis and pathway enrichment analysis to better understand their biological significance using QIAGEN Ingenuity Pathway Analysis (IPA, QIAGEN, Redwood City, CA).

Internal cross-validation of significant metabolites was performed using LOOCV approach, in which each observation is considered as the validation set and the rest ( $N - 1$ ) observations are considered as the training set. The model was built on all the data points (metabolite measurements) except one. The left-out data point was then tested with the LogitBoost method using the model built earlier and the test error associated with the prediction was recorded. The process was repeated for all data points and overall prediction error was computed by taking the average of all these test error estimates recorded for each iteration. This cross-validation was conducted using R package Caret (short for

Classification And REgression Training, version 6.0.88)<sup>42</sup>, and the performance was evaluated using confusionMatrix function to calculate prediction metrics including sensitivity, specificity, and overall accuracy. Multiple methods were applied to extract the optimal predictors as model input to compare validation performance: (1) all significantly abundant metabolites; (2) significantly abundant metabolites with linear dependencies removed; (3) significantly abundant metabolites transformed using PCA to a smaller sub-space where the new PCA variables are uncorrelated with one another.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Raw metabolomics data and analyses related to this study are included in the figures, table, and Supplementary Data 1–5.

### CODE AVAILABILITY

All software versions are indicated within the Methods. All custom scripts related to this study are available from the corresponding author upon reasonable request.

Received: 11 September 2021; Accepted: 16 December 2021;  
Published online: 03 March 2022

### REFERENCES

- Hodgson, S. V., Foulkes, W. D., Eng, C. & Maher, E. R. *A Practical Guide to Human Cancer Genetics*. 4th edn, (Springer, 2014).
- Parenti, I., Rabaneda, L. G., Schoen, H. & Novarino, G. Neurodevelopmental disorders: from genetics to functional pathways. *Trends Neurosci.* **43**, 608–621 (2020).
- Savatt, J. M. & Myers, S. M. Genetic testing in neurodevelopmental disorders. *Front. Pediatr.* **9**, 526779 (2021).
- Yehia, L. & Eng, C. Largescale population genomics versus deep phenotyping: Brute force or elegant pragmatism towards precision medicine. *NPJ Genom. Med.* **4**, 6 (2019).
- Yehia, L. & Eng, C. PTEN Hamartoma Tumor Syndrome. 2001 Nov 29 [Updated 2021 Feb 11]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2022. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1488/>
- Yehia, L., Keel, E. & Eng, C. The clinical spectrum of PTEN mutations. *Annu. Rev. Med.* **71**, 103–116 (2020).
- Liaw, D. et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat. Genet.* **16**, 64–67 (1997).
- Butler, M. G. et al. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J. Med. Genet.* **42**, 318–321 (2005).
- Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584 e523 (2020).
- Yehia, L., Ngeow, J. & Eng, C. PTENopathies: from biological insights to evidence-based precision medicine. *J. Clin. Investig.* **129**, 452–464 (2019).
- Ni, Y. et al. Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *Am. J. Hum. Genet.* **83**, 261–268 (2008).
- Ni, Y. et al. Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. *Hum. Mol. Genet.* **21**, 300–310 (2012).
- Yehia, L. et al. Copy number variation and clinical outcomes in patients with germline PTEN mutations. *JAMA Netw. Open* **3**, e1920415 (2020).
- Clish, C. B. Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb. Mol. Case Stud.* **1**, a000588 (2015).
- Wang, H. et al. Potential serum biomarkers from a metabolomics study of autism. *J. Psychiatry Neurosci.* **41**, 27–37 (2016).
- Yehia, L. et al. Distinct alterations in tricarboxylic acid cycle metabolites associate with cancer and autism phenotypes in Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome. *Am. J. Hum. Genet.* **105**, 813–821 (2019).
- Smith, A. M. et al. A metabolomics approach to screening for autism risk in the children’s autism metabolome project. *Autism Res.* **13**, 1270–1285 (2020).



18. Kang, D. W. et al. Distinct fecal and plasma metabolites in children with autism spectrum disorders and their modulation after microbiota transfer therapy. *mSphere* **5**, <https://doi.org/10.1128/mSphere.00314-20> (2020).
19. Ritz, B. et al. Untargeted metabolomics screen of mid-pregnancy maternal serum and autism in offspring. *Autism Res.* **13**, 1258–1269 (2020).
20. Xu, X. J. et al. Comparison of the metabolic profiles in the plasma and urine samples between autistic and typically developing boys: a preliminary study. *Front. Psychiatry* **12**, 657105 (2021).
21. Needham, B. D. et al. Plasma and fecal metabolite profiles in autism spectrum disorder. *Biol. Psychiatry* **89**, 451–462 (2021).
22. Schmidt, D. R. et al. Metabolomics in cancer research and emerging applications in clinical oncology. *CA Cancer J. Clin.* **71**, 333–358 (2021).
23. Smith, A. M. et al. Amino acid dysregulation metabolotypes: potential biomarkers for diagnosis and individualized treatment for subtypes of autism spectrum disorder. *Biol. Psychiatry* **85**, 345–354 (2019).
24. Ngeow, J. et al. Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-like syndrome characterized by germline PTEN, SDH, or KLLN alterations. *J. Clin. Endocrinol. Metab.* **96**, E2063–E2071 (2011).
25. Tan, M. H. et al. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin. Cancer Res.* **18**, 400–407 (2012).
26. Petersen, A. K. et al. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinform.* **13**, 120 (2012).
27. Hobert, J. A., Mester, J. L., Moline, J. & Eng, C. Elevated plasma succinate in PTEN, SDHB, and SDHD mutation-positive individuals. *Genet. Med.* **14**, 616–619 (2012).
28. Hobert, J. A., Embacher, R., Mester, J. L., Frazier, T. W. 2nd & Eng, C. Biochemical screening and PTEN mutation analysis in individuals with autism spectrum disorders and macrocephaly. *Eur. J. Hum. Genet.* **22**, 273–276 (2014).
29. Meng, W., Huan, Y. & Gao, Y. Urinary proteome profiling for children with autism using data-independent acquisition proteomics. *Transl. Pediatr.* **10**, 1765–1778 (2021).
30. Orozco, J. S., Hertz-Picciotto, I., Abbeduto, L. & Slupsky, C. M. Metabolomics analysis of children with autism, idiopathic-developmental delays, and Down syndrome. *Transl. Psychiatry* **9**, 243 (2019).
31. Kurochkin, I. et al. Metabolome signature of autism in the human prefrontal cortex. *Commun. Biol.* **2**, 234 (2019).
32. Crawley, J. N., Heyer, W. D. & LaSalle, J. M. Autism and cancer share risk genes, pathways, and drug targets. *Trends Genet.* **32**, 139–146 (2016).
33. Fores-Martos, J. et al. Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway abnormalities with cancer. *Mol. Autism* **10**, 17 (2019).
34. Ngeow, J., Stanuch, K., Mester, J. L., Barnholtz-Sloan, J. S. & Eng, C. Second malignant neoplasms in patients with Cowden syndrome with underlying germline PTEN mutations. *J. Clin. Oncol.* <https://doi.org/10.1200/JCO.2013.53.6656> (2014).
35. Mester, J. L. et al. Gene-specific criteria for PTEN variant curation: recommendations from the ClinGen PTEN expert panel. *Hum. Mutat.* **39**, 1581–1592 (2018).
36. Teresi, R. E., Zbuk, K. M., Pezzolesi, M. G., Waite, K. A. & Eng, C. Cowden syndrome-affected patients with PTEN promoter mutations demonstrate abnormal protein translation. *Am. J. Hum. Genet.* **81**, 756–767 (2007).
37. Tan, M. H. et al. A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *Am. J. Hum. Genet.* **88**, 42–56 (2011).
38. Wang, Y. et al. Differential regulation of PTEN expression by androgen receptor in prostate and breast cancers. *Oncogene* **30**, 4327–4338 (2011).
39. Bridgewater Br, E. A. M. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *J. Postgenom. Drug Biomark. Dev.* **04**, <https://doi.org/10.4172/2153-0769.1000132> (2014).
40. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
41. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946–963 (2016).
42. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26, <https://doi.org/10.18637/jss.v028.i05> (2008).

## ACKNOWLEDGEMENTS

We are grateful to our patients and families who contributed to this study. We thank the Genomic Medicine Biorepository of the Cleveland Clinic Genomic Medicine Institute, and our database and clinical research teams. This work was supported in part by the Ambrose Monell Foundation, Breast Cancer Research Foundation, National Cancer Institute (P01CA124570, R01CA118989), the Zacconi Program of PTEN Research Excellence, American Cancer Society (RPG-02-151-01-CCE, Clinical Research Professorship); Doris Duke Distinguished Clinical Scientist Award (all to C.E.); and grant U54NS092090 from the National Institute of Neurological Disorders and Stroke (to C.E. and T.W.F.). L.Y. is an Ambrose Monell Foundation Cancer Genomic Medicine Fellow at the Cleveland Clinic Genomic Medicine Institute. C.E. is the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic and an ACS Clinical Research Professor.

## AUTHOR CONTRIBUTIONS

L.Y., Y.N., and C.E. conceptualized and designed the project. L.Y., Y.N., and T.S. performed the experimental procedures and data analyses. L.Y., Y.N., and C.E. interpreted the data. C.E. supervised the project. All authors drafted, critically revised, and gave final approval of the paper. L.Y. and Y.N. are co-first authors of this study.

## COMPETING INTERESTS

C.E. is an Associate Editor for npj Genomic Medicine but played no role in the review or editorial process. All other authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41525-022-00289-x>.

**Correspondence** and requests for materials should be addressed to Charis Eng.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022