Supplementary Table S1. Characteristics, clinical tasks, and validation maturity of 124 curated records on pediatric microbiome biomarkers.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
1	Mesto, 2026	41621408	United Arab Emirates (Abu Dhabi)	GI/hepatobiliary: Crohn disease	Disease detection	Pediatric CD confirmed by consultant gastroenterologists vs age- and sex-matched children without IBD recruited from the community/pediatric clinics; some controls presented with unrelated acute abdominal pain or constipation. CD cohort included newly diagnosed and longer-duration disease.	n=52 (26 CD, 26 controls). Age groups: CD ≤5 y 2/26, 6-10 y 5/26, >10 y 19/26; controls 4/26, 12/26, 10/26, respectively. Exact mean/median age NR.	Saliva	Full-length (~1, 500 bp) 16S rRNA amplicon sequencing; Oxford Nanopore MinION Mk1C, R9.4 flow cells.	Taxonomic (species-level) and community/diversity.	36 significant species-level candidate biomarkers; 12 composite panels (5-36 species) evaluated using AUC-weighted scores, logistic regression, and LASSO; best method selected by cross-validation.	Internal validation only. Composite-panel AUCs 0.84-1.00; all 36 markers and the 25 species depleted in CD each AUC=1.00; Top-8, Top-10, and 12 good-performer panels AUC=0.96.	No independent/external validation.
2	Shah, 2022 (v2 2023)	36704050	United States and Canada (28 RISK sites; Northeast, Southeast, and West regions)	GI/hepatobiliary: Crohn disease	Differential diagnosis	Treatment-naive pediatric CD vs non-IBD children from the RISK diagnostic cohort; non-IBD controls had gastrointestinal symptoms but did not meet IBD criteria. Analysis focused on inflammatory CD phenotype.	Original RISK cohort n=668 (447 CD, 221 non-IBD controls), age 3-17 y; 1, 321 samples (630 ileal, 387 rectal, 304 fecal). After ileal rarefaction: n=392 (227 CD, 165 controls), mean age 12.2 vs 12.1 y.	Terminal ileal biopsy; rectal biopsy; stool	16S rRNA V4 amplicon sequencing; Illumina MiSeq; secondary reanalysis of RISK data.	Taxonomic (OTU/family/genus-level) microbial profiles.	Random-forest classifier trained in the Northeast region and tested in Southeast and West regions; models built separately for ileal, rectal, and fecal matrices.	Northeast training-region AUROC: ileum 0.89; rectum 0.87; stool 0.82.	Geographic holdout within the same RISK parent cohort (not a new independent cohort): Southeast AUROC ileum/rectum/stool 0.85/0.83/0.85; West 0.91/0.76/0.74.
3	El Mouzan, 2018	30356965	Saudi Arabia (Riyadh)	GI/hepatobiliary: Crohn disease	Differential diagnosis	Children referred for suspected IBD; confirmed CD by standard guidelines vs children in whom CD/IBD was excluded. Common control diagnoses included functional abdominal pain and polyps; controls had no inflammation or infection.	n=35 (15 CD, 20 controls); 78 samples (58 from CD, 20 from controls). Mean age: CD 13.9 y (range 6.7-17.8); controls 13.9 y (3.25-18.6).	Stool; ileal/colonic mucosal forceps biopsies	Fungal ITS amplicon sequencing (ITS1F/ITS2R); bTEFAP; Illumina MiSeq.	Taxonomic mycobiome.	Fungal dysbiosis score using linear/sparse logistic regression on fungal taxa; repeated samples from the same child were averaged before modeling.	Repeated internal validation using 100 random stratified 4/5-training and 1/5-test partitions: stool AUC=0.85 ± 0.057; mucosa AUC=0.71 ± 0.067; P<0.001.	No independent/external validation.
4	El Mouzan, 2021	33642351	Saudi Arabia (Riyadh)	GI/hepatobiliary: Crohn disease	Differential diagnosis	Children presenting with symptoms compatible with CD; confirmed CD according to ESPGHAN guidance vs non-IBD controls without CD/UC/IBD-U, infection, or inflammatory disease. Main control diagnoses: recurrent abdominal pain and polyps.	n=35 (17 CD, 18 controls). Mucosal samples: 44 CD and 14 control; stool: 10 CD and 10 control. Age conflict: CD median 13.9 y in Abstract vs 15.0 y in Results; control median 16.3 y in both.	Stool; ileal and colonic mucosal tissue	Bacterial 16S rRNA V4 amplicon sequencing (515F/806R); bTEFAP; Illumina MiSeq.	Taxon-abundance-derived bacterial dysbiosis score.	Bacterial dysbiosis score using sparse logistic regression on taxon abundances.	5-fold stratified cross-validation with 100 partitions: stool AUC=0.97 ± 0.029; tissue AUC=0.83 ± 0.056 in Results (Abstract reports ±0.055); P<0.001.	No independent/external validation.
5	Shen, 2025	40052570	Public multi-cohort secondary analysis; geographic sources	GI/hepatobiliary: Crohn disease	Disease detection	Public fecal shotgun-metagenomic cohorts. Pediatric CD cohorts explicitly pediatric or	Discovery n=534 (430 CD, 104 controls).	Stool (secondary analysis)	Shotgun metagenomics; raw reads reprocessed through a uniform	Taxonomic species; functional KO genes/pathways; microbial	Random forest after multidimensional feature selection:	Stratified 10-fold cross-validation with Bayesian RF tuning. AUC:	Independent D3 cohort: InDel model AUC=0.996;

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
			heterogeneous			including CD age <14 y; remission and recent antibiotic exposure were excluded. D1+D2 formed discovery; D3 was an independent test cohort. Five other pediatric disease cohorts were used for disease-specificity stress testing.	External D3 n=40 (7 CD, 33 controls). Total analyzed across pediatric CD, adult CD, external, and disease-specificity cohorts=1, 175 metagenomes. Exact pediatric ages/geographic recruitment are NR in the source report.		pipeline.	genomic variants (SVs, InDels, SNVs).	14 species, 16 KO genes, 18 SVs, 19 InDels, 13 SNVs; combined model 27 features. Primary model: 19 microbial InDels.	species 0.959; KO 0.969; SV 0.728; InDel 0.982; SNV 0.869; combined 0.986. InDel model MCC=0.855 and F1=0.972. Upstream feature-selection nesting within CV was not clearly reported.	combined model AUC=0.874. Disease-specificity stress tests against UC, ASD, CeD, NAFLD, and obesity kept InDel AUC around 0.95 after adding non-CD samples.
6	Wang, 2016	26804920	United States (PIBD-CC); RISK cohort	GI/hepatobiliary: Crohn disease	Differential diagnosis	Newly diagnosed, treatment-naive CD vs non-IBD controls. PIBD-CC comprised children undergoing endoscopic IBD evaluation at 7 US centers; non-IBD diagnoses included polyps, IBS, GERD, eosinophilic colitis, H. pylori disease, constipation, and others. RISK was an independent treatment-naive cohort.	PIBD-CC n=87 (24 CD, 63 non-IBD controls), age 1-17 y; mean age 12.6 ± 2.2 vs 11.8 ± 3.9 y. RISK analytical counts reported as 441 ileal samples (187 controls, and 254 CD).	Terminal ileal mucosal biopsy; stool	16S rRNA amplicon sequencing. PIBD-CC: 28F-519R, Roche 454 FLX Titanium. RISK: 16S data reanalysis; original laboratory protocol referenced to prior RISK publication.	Taxonomic/community dysbiosis signatures.	Maximal mutual-information component analysis (MMICA) dysbiosis index; SVM classifiers using disease-associated genera; combined ileal+stool SVM in paired subset.	RISK internal: ileal MMIC AUC=0.84 (5-fold CV); stool SVM AUC=0.72 (95% CI 0.663-0.770). Paired n=74: stool AUC=0.84 (0.747-0.911), ileum 0.81 (0.654-0.925), combined stool+ileum 0.94 (0.908-0.979).	Independent PIBD-CC cohort: first MMIC trained on RISK was applied to independently collected/processed PIBD-CC data, AUC=0.71.
7	Douglas, 2018	29335008	Scotland, UK (BISCUIT); RISK validation dataset	GI/hepatobiliary: Crohn disease	Disease detection; Prognosis/treatment-response prediction	Treatment-naive pediatric CD vs normal-colon controls. CD biopsies were obtained at diagnostic endoscopy before therapy; no systemic antibiotics/steroids in prior 3 months. Treatment-response task used baseline microbiome to classify sustained response vs need for repeat induction/treatment escalation; the endpoint is described as before 24 weeks in Methods but within 150 days in Results. Induction therapies were heterogeneous.	Initial BISCUIT n=40 (20 CD, 20 controls), all <17 y, mean age 12.7 y. Two CD microbial profiles were excluded because of low sequencing depth, leaving 38 participants (18 CD, 20 controls) for microbial disease-state models and 18 CD patients for microbial treatment-response models. Figure 1 nevertheless labels the initial MGS dataset as n=40, while stating	Intestinal biopsies	16S rRNA V6–V8 amplicon sequencing (Illumina MiSeq, PE300) plus shotgun metagenomics (Illumina HiSeq, PE100); host genotypes were also recovered from MGS reads.	Taxonomic; functional KO/pathway/module; host genetic-risk features (multi-omics comparison).	Random-forest models across 19 taxonomic/functional datasets with LOOCV; 16S genus-level models were strongest for disease state and treatment response; combined top-feature RFs were also evaluated.	Disease state: best 16S genus accuracy=84.2% (P<0.001); best MGS strain=68.4% (P=0.016); combined=78.9% (P<0.001). Treatment response: 16S genus=77.8% (P=0.008); MGS strain/genus/pathway each=72.2%; combined=94.4% (P<0.001).	Independent RISK data were used to retrain a new RF on selected BISCUIT-derived 16S features: accuracy=73.2% (P<0.001), with changed feature rankings. This is model redevelopment, not direct locked-model validation. No independent treatment-response validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							that the two CD samples were excluded from downstream microbial MGS analyses; host genetic profiles remained available for all 40 participants.						
8	Bosch, 2018	29788410	Netherlands	GI/hepatobiliary: IBD vs IBS	Disease detection; Differential diagnosis	Rome IV IBS/FAP-NOS vs de novo IBD (15 CD, 15 UC) and age/sex-matched healthy controls. The primary clinical comparison was IBS/FAP-NOS vs IBD.	n=75: IBS/FAP-NOS 15 (9 IBS/6 FAP-NOS), IBD 30 (15 CD/15 UC), healthy controls 30. Median age [IQR] (range), y: CD 12.8 [5.0] (5.9-17.9); UC 11.8 [7.8] (3.2-17.8); IBS/FAP-NOS 12.9 [8.4] (4.4-18.1); controls 12.7 [8.1] (4.1-17.9).	Stool (fecal headspace VOCs)	Field asymmetric ion mobility spectrometry (FAIMS; Lonestar, Owlstone Ltd.); each sample measured in triplicate, third matrix analyzed.	Metabolites/VOC pattern (microbiota-related indirect signature).	44 wavelet-derived VOC features; sparse logistic regression, random forest, Gaussian process, and SVM; best classifier reported per comparison.	10-fold CV with feature selection and training confined to the 90% training fold. IBS/FAP-NOS vs IBD: AUC=0.94 (95% CI 0.88-1.00), sensitivity 1.00, specificity 0.87, PPV 0.79, NPV 1.00. IBD vs healthy controls: AUC=0.96 (0.93–1.00). IBS/FAP-NOS vs healthy controls: AUC=0.59 (0.41-0.77); CD vs UC: AUC=0.67 (0.47-0.88).	No independent/external validation.
9	Prendergast, 2026	42577488	New Zealand (Christchurch/Canterbury)	GI/hepatobiliary: Celiac disease	Disease detection; Treatment-response monitoring	Children evaluated for suspected CeD at Christchurch Hospital. Untreated CeD required biopsy confirmation; controls were clinical endoscopy comparators with negative CeD serology and normal duodenal biopsy. Nine CeD children were resampled after 6 months of gluten-free diet.	21 unique baseline participants (11 untreated CeD, 10 controls); 9 of 11 CeD resampled at 6 months, yielding 30 sample-timepoints. Age mean ± SD: untreated 10.7 ± 2.9 y; treated repeat samples 10.7 ± 2.9; controls 10.2 ± 2.8.	Stool	16S rRNA V3-V4 amplicon sequencing (Illumina MiSeq, 2x250 bp) plus stool 1H NMR and targeted GC-MS metabolomics.	Multi-omics stool microbiome–metabolome signature: species-level bacterial taxonomic features plus NMR/GC-MS metabolites.	Late-fusion elastic-net classifier combining species-level microbiome and metabolome probabilities; inner 3-fold tuning and stratified outer 4-fold CV repeated 5, 000 times. DIABLO cross-omic modeling was exploratory.	Cross-validated performance (pooled untreated+treated CeD vs controls): microbiome ROC-AUC median 0.659 (IQR 0.558-0.729), PR-AUC 0.791; metabolome 0.612/0.733; fused 0.682 (0.612-0.741)/0.776. Fused sensitivity 0.49, specificity 0.68, balanced accuracy 0.59.	No independent/external validation.
10	El Mouzan, 2023	37155522	Saudi Arabia (Riyadh)	GI/hepatobiliary: Celiac disease	Disease detection; Differential diagnosis	Children with new-onset CeD confirmed according to standard diagnostic criteria and without antibiotic exposure for ≥6 months. Controls comprised 20 clinically healthy schoolchildren providing stool samples and 19	CeD n=40; non-CeD controls n=39 (20 stool healthy controls, 19 duodenal clinical controls).	Stool; second-part duodenal mucosal tissue	Shotgun metagenomic sequencing on HiSeq; bacteria, viruses, and fungi profiled.	Multi-kingdom taxonomic/community profiles.	Whole-community ROC analysis; Boruta random-forest importance analysis. Confirmed candidate taxa: fecal Bacteroides intestinalis and Burkholderiales bacterium 1-1-47;	No resampling/holdout procedure was reported for the ROC analysis. Stool AUC: bacteria 0.52, viruses 0.58, fungi 0.677, bacteria+viruses 0.818. Mucosal AUC: bacteria	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						children undergoing diagnostic endoscopy for clinical indications who provided second-part duodenal tissue; the latter had normal CeD serology, endoscopy, and duodenal histopathology.	Matrix-specific CeD sample counts and age were NR in this report.				mucosal Human endogenous retrovirus K.	0.812, viruses 0.586, fungi 0.35.	
11	Zafeiropoulou, 2020	32791131	Scotland, UK (Glasgow)	GI/hepatobiliary: Celiac disease	Disease detection	Biopsy-confirmed untreated/new-onset CeD on a gluten-containing diet, treated CeD on GFD, healthy controls, and unaffected siblings. Thirteen new-onset children provided complete 6- and 12-month follow-up after GFD initiation.	Baseline n=141: 20 untreated CeD, 45 treated CeD, 57 healthy controls, 19 unaffected siblings; total 167 stool samples. Complete longitudinal subset n=13. Age mean (SEM), y: HC 7.8 (0.41), siblings 9.1 (0.76), untreated CeD 10.1 (0.70), treated CeD 9.3 (0.47).	Stool	16S rRNA V4 amplicon sequencing (Illumina MiSeq); fecal SCFAs/BCFAs measured by gas chromatography and other targeted assays.	Taxonomic/community signature; metabolite measurements were supportive rather than part of the diagnostic classifier.	11 OTUs consistently lower in CeD irrespective of GFD; random-forest classifier with 10, 000 trees.	RF internal evaluation: OOB error=21.5%; ROC-AUC=0.789; permutation-based P<0.001. The 11-OTU panel was selected using the same overall cohort before RF modeling; nested feature selection was not reported, so performance may be optimistic. The manuscript does not explicitly state whether the ROC-AUC was derived from OOB predictions. Longitudinal GFD sampling assessed temporal microbiome change rather than treatment-response prediction.	No independent/external validation.
12	Schwimmer, 2019	31255652	United States (San Diego, California)	GI/hepatobiliary: Pediatric NAFLD/MASLD	Disease detection; Phenotyping/stratification	Biopsy-proven pediatric NAFLD diagnosed within 90 days vs overweight/obese children without liver disease (history, laboratory testing, and liver MRS). Within NAFLD, histology defined NASH and fibrosis severity.	n=124 (87 NAFLD, 37 controls), age 8-17 y. Median age [IQR]: cases 12 [10, 14] y; controls 13 [11, 14] y.	Stool	16S rRNA V1-V3 amplicon sequencing (Illumina MiSeq, 300-bp paired-end) plus metagenomic whole-genome shotgun sequencing (Illumina HiSeq 2000).	Taxonomic/community and functional KO/pathway features; clinical-microbiome composite.	CART models combining ALT with microbial functional genes/pathways: LPS-biosynthesis KOs (e.g., kdsC/lpxC) and flagellar-assembly KOs (e.g., FlhA).	CART models used 10% leave-out cross-validation. AUROC: LPS-based NAFLD vs obese controls 0.95; definite NASH vs NAFLD-not-NASH 0.92; flagellar-based NAFLD vs controls 0.97; moderate/severe (stage ≥2) vs absent/mild fibrosis 0.87.	No independent/external validation.
13	Lin, 2025	40091070	Taiwan	GI/hepatobiliary: Pediatric NAFLD/MASLD	Disease detection	Obese children (BMI >95th percentile) with MASLD vs age-, sex-, and BMI-matched obese children without MASLD; MASLD required steatosis plus cardiometabolic risk factors.	n=52 (26 MASLD, 26 matched non-MASLD controls), age 6-18 y. Source Table 1 reports mean ± SEM: non-MASLD 12.7 ± 2.3 y; MASLD 12.5 ± 2.1 y.	Stool	Short-read V3-V4 16S rRNA sequencing (Illumina MiSeq) and full-length V1-V9 16S rRNA sequencing (PacBio Sequel IIe, CCS).	Taxonomic ASV/genus/species features.	Two-step random-forest models: initial top-35-abundance or LDA>3 taxa, then top 20% by MDA; final FL16S model 24 features and V3-V4 model 10 features.	10-fold cross-validation. FL16S AUC=86.98% vs V3-V4 AUC=70.27% (DeLong P=0.0079). The authors state that feature selection used the entire dataset before CV, so optimism from non-nested selection cannot be excluded.	No independent/external validation.
14	Zhang, 2024	38189294	China	GI/hepatobiliary: Pediatric NAFLD/MASLD	Disease detection	Children with NAFLD (BMI >95th percentile plus clinical evaluation/liver ultrasonography) vs healthy controls; recent	n=114 (79 NAFLD, 35 controls), age 7-18 y. Age: controls 11.42 ±	Stool	16S rRNA V3-V4 amplicon sequencing (Illumina MiSeq); shotgun metagenomics (Illumina	Taxonomic bacterial panel; functional/metagenomic pathways used for	10-taxon 16S random-forest panel (Sneathia, Lachnospiraceae ND3007 group, Prevotella,	Performance reporting is internally inconsistent: Methods state a 2/3 training–1/3 test split but also state	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						antibiotics/probiotics excluded. Thirteen NAFLD children had liver biopsy; 15 were reassessed after 3-6 months of lifestyle treatment for longitudinal microbiome/BMI associations.	1.08 y; NAFLD 12.23 ± 2.63 y (meaning of ± not specified in the table).		NovaSeq 6000) in the 79 NAFLD samples.	complementary analyses.	Lachnoclostridium, Escherichia-Shigella, Prevotellaceae UCG-001, Haemophilus, Lactobacillus, Ruminiclostridium 6, Coprococcus 2).	that the model was trained on all 114 samples; the origin of ROC predictions is therefore unclear. Reported AUC=0.969, sensitivity=0.914, specificity=0.949.	
15	de Meij, 2016	27760208	Netherlands (Amsterdam)	GI/hepatobiliary: Functional constipation	Disease detection	Refractory functional constipation diagnosed by Rome III at tertiary centers vs healthy Dutch children from a prior longitudinal microbiota cohort. Laxatives were stopped for 4 weeks before sampling; major GI/neurologic disorders and recent antibiotics/steroids/immunosuppression were excluded.	n=137 (76 constipation, 61 healthy controls). Median age: 8.0 y (range 4.2-17.8) vs 8.6 y (4.1-17.9).	Stool	IS-pro PCR-based profiling of species-specific 16S-23S rDNA interspace length plus phylum-specific 16S sequence polymorphisms; ABI Prism 3130XL fragment analysis.	Taxonomic fingerprint/community profile.	Adaptive group-regularized logistic ridge regression; most discriminative taxa included Bacteroides fragilis, B. ovatus, Bifidobacterium longum, Parabacteroides spp. (higher) and Alistipes finegoldii (lower).	10-fold cross-validation: AUC=0.78 ± 0.08 (paper labels this as AUC ± 95% CI), sensitivity=90%, specificity=73%, accuracy=82%.	No independent/external validation.
16	Iwasawa, 2018	29615776	Japan (Kanagawa)	GI/hepatobiliary: Primary sclerosing cholangitis	Disease detection; Differential diagnosis	Pediatric-onset PSC vs age-matched UC without PSC and healthy controls. PSC diagnosis used cholangiographic bile-duct changes plus clinical/biochemical criteria; UC used revised Porto criteria. Medication effects were examined in sensitivity analyses.	n=64 (24 PSC, 16 UC, 24 healthy controls). Median age: PSC 12.5 y; UC 12.5 y; controls 11.5 y; pediatric onset <18 y.	Saliva	16S rRNA V1-V2 amplicon sequencing; Roche 454 GS FLX Titanium.	Taxonomic genus/species oral microbiome and community structure.	AUC-RF random-forest models; best genus panels used 8 genera for PSC vs HC and 6 genera for PSC vs UC; best species panels used 16 and 11 species, respectively.	10-fold cross-validation repeated 20 times. Genus-level AUC: PSC vs HC 0.7423; PSC vs UC 0.8756. Species-level AUC: PSC vs HC 0.8011; PSC vs UC 0.7626.	No independent/external validation.
17	Wang, 2026	41885920	China (Zhoushan, Zhejiang)	GI/hepatobiliary: Feeding intolerance	Prognosis/treatment-response prediction	Preterm infants admitted to a tertiary maternal/child hospital; feeding-tolerant (FT) vs feeding-intolerant (FI) groups defined from day-21 feeding assessments in the electronic record. FI criteria included gastric residuals, frequent vomiting/abdominal distension, or interruption of enteral feeding.	n=49 (34 FT, 15 FI). Gestational age, mean ± SD: FT 33.21 ± 1.67 wk; FI 32.88 ± 1.85 wk. Birth weight: 1972.35 ± 391.14 vs 1766.00 ± 435.71 g. Stool collected at postnatal day 5-7.	Stool	16S rRNA V3-V4 amplicon sequencing, Illumina MiSeq PE300; untargeted fecal metabolomics by UHPLC coupled to Orbitrap Exploris 120 MS.	Multi-omics: taxonomic bacterial features + fecal metabolites.	Candidate genera from differential analysis/logistic regression plus random-forest-ranked metabolites; reported individual markers included Clostridium_P, Escherichia, Staphylococcus, Bifidobacterium, Arg-Pro, Glu-Arg, and lactaldehyde. No fixed combined panel performance was reported.	Apparent ROC performance in the same n=49 cohort: Clostridium_P AUC=0.947; Bifidobacterium 0.851; Escherichia 0.765; Staphylococcus 0.733; Arg-Pro 0.920; Glu-Arg 0.873; lactaldehyde 0.900. No held-out diagnostic test set reported.	No independent/external validation.
18	Liu, 2022	35252022	China (Chongqing)	GI/hepatobiliary: Feeding intolerance	Disease detection	Prospective VLBW infants with gestational age ≤30 wk and birth weight <1500 g; FI required all three criteria: gastric residual >50% of previous feed, emesis/abdominal distension, and decreased/delayed/discontinued enteral feeding; compared with feeding-tolerant infants. Temporal relation of the 2-wk sample	n=27 (14 FI, 13 FT); 49 fecal samples at 2 and 4 wk. Gestational age, median [IQR]: FI 28.93 [27.61-29.50] wk; FT 29.14 [28.07-29.43] wk. Birth weight, mean ±	Stool (2 and 4 wk)	16S rRNA V3-V4 amplicon sequencing (338F/806R), Illumina MiSeq 2x250; fecal SCFAs by GC-MS; AI-2 measured by Vibrio harveyi BB170 bioluminescence assay.	Taxonomic/community features + microbial metabolites/signaling (SCFAs, AI-2).	No multivariable microbiome classifier. ROC analyses evaluated individual SCFAs and AI-2 at 2 wk; longitudinal community changes were also assessed.	At 2 wk: propanoic acid AUC=0.878; valeric acid AUC=0.816; hexanoic acid AUC=0.744 in Results/Fig. 6; AI-2 AUC=0.736. Abstract instead labels the three SCFAs as propanoic, butanoic, and valeric acids (internal inconsistency).	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						to FI onset was unclear.	SD: FI 1110.14 ± 199.943 g; FT 1173.85 ± 177.132 g.						
19	Santucci, 2026	42444523	United States (Cincinnati, Ohio)	GI/hepatobiliary: Acute pancreatitis / Crohn disease / IBS	Differential diagnosis	Previously established pediatric/young-person cohorts reprocessed uniformly: acute pancreatitis (AP), active Crohn disease (CD), Rome IV IBS, and healthy controls (HC). Only baseline samples were analyzed; AP samples were obtained within 7 days of symptom onset.	n=120: AP 30, CD 29, IBS 27, HC 34. Overall age median [IQR] 15.2 [12.4-17.1] y.	Stool	Shotgun metagenomic sequencing; taxonomic profiles with sylph and functional profiles with HUMAnN3/MetaCyc.	Taxonomic species abundance; functional pathways were characterized secondarily.	Random-forest classifier using species relative abundance; top 10 species ranked by variable importance.	Out-of-sample multiclass AUC=0.90, specificity=90%; one-vs-rest AUC: CD 0.97 (95% CI 0.96-0.98), AP 0.92 (0.89-0.95), HC 0.88 (0.86-0.91), IBS 0.83 (0.79-0.87). Exact resampling/holdout procedure is NR in the source report.	No independent external model validation: cohorts came from prior studies at the same tertiary institution and were reprocessed together.
20	Guo, 2018	30483228	China (Hebei Province)	GI/hepatobiliary: Infant cholestasis	Disease detection	Infants with cholestasis (including CMV hepatitis, biliary atresia, and other/undetermined causes) vs healthy infants undergoing physical examination. Cholestasis was defined by abnormal bilirubin/GGT criteria.	n=80 (43 infantile cholestasis, 37 healthy). Age, mean ± SD: 2.25 ± 1.79 vs 12.03 ± 8.75 months, respectively.	Stool	16S rRNA V3-V4 amplicon sequencing; Illumina MiSeq 300-nt paired reads; PICRUSt/KEGG functional prediction.	Taxonomic/community features; predicted functional categories.	Random forest on genus composition; 28 candidate biomarkers, with Rothia, Eggerthella, Phyllobacterium, and Blautia among the top features.	Samples randomized into training and testing sets; test set used for biomarker validation. ROC AUC >0.97 with five repeats. Exact split proportion and whether feature selection was fully restricted to training data were NR.	No independent/external cohort. The test set was a random split from the same study sample.
21	Wang, 2020	32352720	China (Shanghai)	GI/hepatobiliary: Chronic pancreatitis	Disease detection; Phenotyping/stratification	Children with chronic pancreatitis diagnosed by INSPPIRE criteria vs healthy controls, ages 4-18 y; recent antibiotics/probiotics excluded. Host functional-gene mutations were evaluated in a subset of cases.	n=65 (30 chronic pancreatitis, 35 healthy controls). Mean age: 7.2 ± 0.5 vs 8.3 ± 0.7 y. Peripheral-blood mutation data were available for 21 cases.	Stool	16S rRNA V3-V4 amplicon sequencing (338F/806R), Illumina MiSeq; predicted microbial functional pathways; host-gene mutation testing in cases.	Taxonomic/community features; predicted functions; host-genotype-associated microbial stratification.	Six-genus panel (Faecalibacterium, Subdoligranulum, Phascolarctobacterium, Bifidobacterium, Eubacterium, Collinsella) combined by multivariable logistic regression.	Same-cohort apparent performance: six-genus model AUC=0.92 (reported 95% CI 0.84-0.92), P<0.001. Individual genus AUCs approximately 0.65-0.80. No resampling/held-out test set was reported. Host-mutation subgroup analyses were association-only; no subgroup classification performance was reported.	No independent/external validation.
22	Li, 2021	34481999	Canada (Ottawa, Ontario)	GI/hepatobiliary: Ulcerative colitis	Disease detection	Children <18 y undergoing diagnostic colonoscopy for suspected UC; new-onset UC vs non-UC controls with visually and histologically normal mucosa.	n=73 (42 UC, 31 controls). Mean ± SD age: UC 13.7 ± 4.0 y; controls 13.1 ± 3.5 y. Table 1 medians [range]: 14.2 [5.1-17.5] vs 15.0 [3.5-17.7] y.	Ascending-colon mucosal-luminal interface aspirate	Glycomics: MALDI-TOF-MS (4800 MALDI-TOF/TOF) and porous-graphitized-carbon nano-LC-MS (Eksigent nano-LC coupled to Q Exactive HFX).	Metabolites/biochemical features: microbiota-associated N-glycans.	PLS-DA on glycan profiles; candidate paucimannosidic/truncated N-glycans. Strongest marker Man2GlcNAc2Fuc1; cross-platform candidate Man3GlcNAc3Fuc1Gal1NeuNAc1.	Same-cohort PLS-DA accuracy=0.91 (MALDI: Q2=0.43, R2=0.57; PGC-LC-MS: Q2=0.67, R2=0.70). Man2GlcNAc2Fuc1 AUC=0.9894 by MALDI and 0.9838 (95% CI 0.9608-1.0) by PGC-LC-MS. No resampling/held-out validation reported.	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
23	El Mouzan, 2025	39523762	Saudi Arabia	GI/hepatobiliary: Ulcerative colitis	Disease detection	New-onset, treatment-naive UC with no antibiotic exposure for at least 6 months vs healthy school children; feces collected before bowel preparation.	n=40 (20 UC, 20 healthy controls). Median age [range]: UC 13 [0.5-21] y; controls 13 [7-16] y. The reported UC range includes participants >18 y and cannot be separated from the source report.	Stool	Shotgun metagenomic sequencing; Nextera XT libraries; Illumina NovaSeq S4, 2x150 bp.	Taxonomic mycobiome (fungal species/genera).	Random-forest-selected fungal signatures: Melampsora pinitorqua, Onygenales_u_s, and Saccharomycodes ludwigii (species-level candidates).	Same-cohort apparent ROC: species-level fungal microbiome AUC=0.976; three RF-selected species AUC=0.984 (95% CI 0.951-1.00). No cross-validation/held-out test set was reported.	No independent/external validation.
24	El Mouzan, 2024	38596448	Saudi Arabia (Riyadh)	GI/hepatobiliary: Ulcerative colitis	Disease detection	New-onset untreated UC recruited from multiple Riyadh hospitals, no antibiotic exposure for at least 6 months, vs randomly selected healthy school children.	n=40 (20 UC, 20 healthy controls). Median age [range]: UC 14 [0.5-21] y; controls 12.9 [6.8-16.3] y. The reported UC range includes participants >18 y and cannot be separated.	Stool	Shotgun metagenomic sequencing; Nextera XT libraries; Illumina NovaSeq S4.	Taxonomic bacterial and bacteriophage profiles.	Random forest identified 11 bacterial species and 4 bacteriophage species as important biomarkers; ROC analyses were performed for all species and RF-selected subsets.	Apparent same-cohort AUC: all bacteria 0.895 (95% CI 0.791-1.000); selected 11 bacteria 0.976 (0.942-1.000); all bacteriophages 0.874 (0.759-0.988); selected 4 bacteriophages 0.945 (0.878-1.000).	No independent/external validation.
25	Zuo, 2022	35440670	United States (Los Angeles, California; primary cohort)	GI/hepatobiliary: Ulcerative colitis	Disease detection	Primary cohort: mild-to-moderate UC vs healthy controls; some UC patients were receiving steroids, biologics, immunomodulators, or 5-ASA. Independent set: 7 pediatric UC cases aged 7–17 y vs 8 healthy controls, with all 8 controls aged ≥21 y.	Primary n=42 (19 UC, 23 healthy), ages 7-21 y (mixed pediatric/young adult). Independent validation n=15 (7 UC age 7-17 y; 8 adult healthy controls ≥21 y).	Stool	Paired profiling on the same samples: 16S rRNA V4 amplicon sequencing (Illumina MiSeq 2x150) and shotgun metagenomics (Illumina NextSeq 500 2x150); species/pathways derived from shotgun data.	Taxonomic genus/species features + functional pathway abundance.	Random forest (500 trees); separate 16S-genus, shotgun-species, and shotgun-pathway models; optional age/sex covariates.	Five-fold cross-validation repeated 50 times. AUROC without covariates: 16S genus 0.908 (95% CI 0.810-1.000); shotgun species 0.910 (0.813-1.000); pathways 0.951 (0.878-1.000). With age/sex, pathway AUROC=0.964 (0.902-1.000).	Independent set not used in training: highest AUROC=0.869 for the 16S model. Important limitation: all 8 validation controls were adults ≥21 y, so this is not a fully pediatric external validation.
26	Li, 2025	40297136	Primary public dataset GSE93624 (original recruitment region NR in the source report); independent validation under Tianyou Hospital IRB, China	GI/hepatobiliary: Crohn disease (primary); pediatric IBD (validation)	Differential diagnosis	Primary secondary analysis: pediatric Crohn disease ileal-biopsy RNA-seq vs age-matched non-IBD ileal-biopsy controls with normal histopathology. Independent validation: pediatric IBD vs non-IBD stool samples.	Primary n=245 pediatric ileal biopsies; exact CD/control split and ages are NR in the source report. Validation n=36 stool samples (19 IBD, 17 non-IBD).	Ileal biopsy (primary); stool (validation)	Primary: RNA-seq non-host reads profiled taxonomically with Kraken2. Validation: 16S rRNA amplicon sequencing with QIIME2/DADA2; detailed region/instrument NR in the source report.	Taxonomic microbial profiles (phylum/genus/species).	Logistic regression, random forest, SVM, and XGBoost with forward feature selection; XGBoost performed best; Orthotospovirus and Vescimonas were highlighted as key genera.	Primary analysis used a 70/30 train/test split with grid-search 5-fold cross-validation for model tuning. At the genus level, ROC-AUCs were 0.79 for logistic regression, 0.62 for random forest, 0.77 for SVM, and 0.82 for XGBoost. At the species level, ROC-AUCs were 0.77, 0.70, 0.76,	Independent n=36 validation cohort reproduced directional differences for Orthotospovirus and Vescimonas (and Actinomyces among traditional candidates), but the same locked

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
												and 0.83, respectively. XGBoost showed the highest discrimination at both taxonomic levels.	classifier was not reported with external performance metrics; biomarker reproducibility only.
27	Vermeer, 2025	40503566	Netherlands (Amsterdam)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection; Differential diagnosis	Development: newly diagnosed treatment-naive IBD vs healthy controls. Validation: independent prospective cohort of newly diagnosed treatment-naive IBD vs children with gastrointestinal symptoms in whom IBD was excluded after guideline-based work-up.	Development n=117 (58 IBD: 28 CD/30 UC; 59 healthy); age median [IQR] 14 [11-15] vs 7 [5-11] y. Validation n=81 (43 IBD: 24 CD/16 UC/3 IBD-U; 38 symptomatic controls); age 15.1 [11.1-16.4] vs 15.0 [12.9-16.6] y.	Stool	Microbiota by 16S-23S rDNA InterSpace Profiling (IS-pro). Amino acids: targeted HPLC in discovery; HILIC-tandem MS in validation.	Multi-omics: taxonomic bacterial features + fecal amino acids.	Six-marker logistic model selected after LASSO/Elastic Net: Escherichia coli, Alistipes finegoldii, leucine, ornithine, taurine, and alpha-aminoadipic acid.	Random train/test assignment; 10-fold cross-validation for hyperparameter tuning; LASSO/Elastic Net feature selection repeated 50 times. Six-marker logistic model AUC=0.94 (95% CI 0.89–0.98) in the Results/Abstract/Discussion; Figure 2 reports 0.96 (0.90–1.00), an unresolved internal inconsistency.	Independent prospective validation against symptomatic non-IBD controls: the same six-marker panel AUC=0.84 (95% CI 0.67–0.95), sensitivity=79%, specificity=87%. Leucine alone AUC=0.89 (0.81–0.95), sensitivity=81%, specificity=81%.
28	Vermeer, 2024	38732837	Netherlands (Amsterdam)	GI/hepatobiliary: Inflammatory bowel disease	Differential diagnosis; Prognosis/treatment-response prediction	De novo therapy-naive pediatric IBD vs controls with gastrointestinal symptoms (CGI). Secondary analysis predicted 3-month response to induction therapy from baseline VOCs; response required clinical response without treatment escalation.	n=184 (109 IBD: 74 CD/29 UC/6 IBD-U; 75 CGI), age 4-17 y. Median age [IQR]: IBD 15.0 [12.5-16.0], CGI 14.0 [11.0-16.0]. Response analysis n=72 (34 responders, 38 nonresponders).	Stool headspace	Gas chromatography-ion mobility spectrometry (GC-IMS), FlavourSpec.	Metabolites/VOCs: untargeted fecal VOC pattern.	Feature selection within training folds; sparse logistic regression and XGBoost classifiers; best-performing algorithm reported for each comparison.	10-fold internal CV with feature selection restricted to training data. IBD vs CGI AUC=0.71 (95% CI 0.64-0.79), sensitivity=0.59, specificity=0.77; CD vs CGI 0.75 (0.67-0.84); UC vs CGI 0.67 (0.56-0.78); CD vs UC 0.57 (0.45-0.69). Baseline response prediction AUC=0.70 (0.58-0.83), P=0.10 (not significant).	No independent/external validation.
29	Wang, 2021	34950604	China (Shanghai)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection; Differential diagnosis	Development: newly diagnosed untreated pediatric IBD vs healthy controls. Validation: independent newly diagnosed early-onset IBD vs Rome IV IBS, a clinically relevant symptomatic comparator.	Development n=93 (66 IBD: 54 CD/8 UC/4 IBD-U; 27 healthy), mean age 10.0 ± 5.3 vs 7.1 ± 3.8 y. Validation n=62 (14 IBD: 12 CD/2 UC; 48 IBS), mean age 2.3 ± 3.1 vs 6.7 ± 3.5 y.	Stool	16S rRNA V3-V4 amplicon sequencing (338F/806R), Illumina MiSeq 2x300; Tax4Fun imputed functional profiles.	Taxonomic OTU-level bacterial features.	Random forest on 1, 902 OTUs; 10-fold CV selected an 11-OTU panel and probability-of-disease (POD) score.	Development IBD vs healthy: AUC=0.88 (95% CI 0.80-0.95).	Independent validation IBD vs IBS: AUC=0.84 (95% CI 0.72-0.96). The validation cohort was markedly younger in IBD than IBS.
30	de Meij, 2018	30102706	Netherlands (Amsterdam)	GI/hepatobiliary: Inflammatory bowel	Disease detection; Prognosis/treatment-respons	Prospective de novo treatment-naive IBD (CD/UC) vs healthy pediatric controls; serial	Baseline cohort n=165: 104 IBD (63	Stool (baseline; 1, 3, 6 wk; subset at 3 mo)	IS-pro PCR-based 16S-23S rDNA interspace profiling.	Taxonomic bacterial fingerprint/community	Group-regularized logistic ridge regression (GRR), age/sex	Age/sex-adjusted GRR; penalty tuning by leave-one-out CV and	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
				disease	e prediction	stool after treatment. In CD, baseline microbiota were also tested for prediction of successful 6-wk exclusive enteral nutrition (EEN).	CD, 41 UC) and 61 healthy controls. Median age: IBD 14.0 y, HC 7.8 y. Classification analytic dataset after preprocessing: n=148. Reported CD age range 5.8–19.0 y conflicts with the stated eligibility criterion of age <18 y. Six-week remission: 44/63 CD and 29/41 UC. EEN prediction dataset n=33 CD.			features.	adjusted; post-hoc variable selection yielded 17 features corresponding to 12 species. Separate GRR assessed baseline prediction of 6-wk EEN success.	predictive performance by 5-fold CV. Post-hoc variable-selected 17-feature/12-species model AUC≈0.87; full ridge/GRR ≈0.90. Baseline prediction of 6-week EEN success was poor: AUC=0.55 with post-hoc selection and ≈0.64 without selection.	
31	van Gaal, 2017	28439048	Netherlands (two tertiary centers, Amsterdam)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection; Phenotyping/stratification	De novo treatment-naive pediatric IBD vs age-matched asymptomatic healthy controls; secondary within-IBD comparison of CD vs UC.	n=60 (36 IBD: 23 CD/13 UC; 24 healthy), age 4-17 y. Median age [IQR as reported]: controls 11.46 [7.46-13.3], CD 14.24 [12.49-16.24], UC 14.16 [9.66-15.29] y.	Stool headspace	Field asymmetric ion mobility spectrometry (FAIMS), Lonestar; 2D wavelet preprocessing.	Metabolites/VOCs: untargeted fecal VOC pattern.	Within each fold, Wilcoxon feature selection retained the two most informative features; RF, Gaussian-process, XGBoost, and sparse logistic regression were compared and the best classifier was reported per task.	Five-fold CV (80% train/20% test per fold). IBD vs healthy AUC=0.76 ± 0.14, sensitivity=79%, specificity=78%; CD vs healthy 0.90 ± 0.10, 83%/83%; UC vs healthy 0.74 ± 0.19, 77%/75%. CD vs UC AUC=0.67 ± 0.19, P=0.0996 (not significant).	No independent/external validation.
32	de Meij, 2014	25248313	Netherlands (Amsterdam)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection; Phenotyping/stratification	Prospective newly diagnosed pediatric IBD (UC/CD) vs controls; repeated VOC assessment after remission at 6 wk. Disease activity assessed by physician global assessment, CRP, and fecal calprotectin.	n=83 (26 UC, 29 CD, 28 controls). Age NR in the conference abstract.	Stool headspace	Electronic nose (Cyranose 320).	Metabolites/VOCs: fecal gas/VOC pattern.	Electronic-nose pattern classifier; internally cross-validated ROC curves. Exact classifier and resampling scheme NR in the abstract.	Active disease: UC vs controls AUC=1.00 ± 0.00, sensitivity/specificity=100%/100%; CD vs controls 0.85 ± 0.05, 86%/67%; CD vs UC 0.96 ± 0.03, 97%/92%. At remission: UC vs controls 0.94 ± 0.06, 94%/94%; CD vs controls 0.94 ± 0.06, 94%/94%; CD vs UC 0.81 ± 0.08, 88%/72%. No direct active-vs-remission classifier was reported.	No independent/external validation. Evidence is conference-abstract level only.
33	Olbjorn, 2022	3588899	Norway (South-Eastern Norway)	GI/hepatobiliary:	Differential diagnosis;	Treatment-naive pediatric IBD from three	n=235: 110 IBD (80	Stool	Targeted absolute-quantification	Taxonomic species/subspecies	Weighted bacterial-abundance	Exploratory same-cohort index	Not evaluated: no

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
		2		Inflammatory bowel disease	Phenotyping/stratification; Prognosis/treatment-response prediction	prospective population-based studies vs symptomatic non-IBD patients with normal endoscopy/histology and MRI, plus healthy controls. CD vs UC and later need for biologic therapy were also modeled.	CD, 27 UC, 3 IBD-U), 50 symptomatic non-IBD, 75 healthy. Median age [range]: CD 13 [0.74-17.9], UC 11.5 [4-17], non-IBD 12 [3.7-18], healthy 10 [2-17.9] y.		qPCR (PMP/OpenArray, QuantStudio 12K), >100 bacterial species/subspecies; analyses used relative abundance.	features.	indexes with logistic regression: Diagnostic Indexes (IBD vs healthy; non-IBD vs healthy; IBD vs non-IBD), Phenotype Index (CD vs UC), and Prognostic Index (future biologic vs conventional therapy).	development. Candidate taxa were selected from observed between-group abundance differences, dichotomized relative to group medians, weighted, and summed into diagnostic/phenotype/prognostic indexes. Logistic regression OR CIs were obtained using 10, 000-repetition bootstrap. LOOCV was performed, but its reported performance metric was MSE rather than cross-validated AUC. Apparent ROC performance: IBD vs healthy AUC=0.78 (95% CI 0.72–0.85); non-IBD vs healthy 0.77 (0.68–0.86); IBD vs symptomatic non-IBD 0.69 (0.60–0.78), improving to 0.83 (0.74–0.93) in a subset tested with 22 additional targets; CD vs UC 0.74 (0.64–0.85); future biologic-therapy index 0.72 (0.63–0.82), OR=2.1 per index unit.	independent/external cohort; all indexes were developed/evaluated in the same study population.
34	Eck, 2017	28330889	Netherlands (Amsterdam; technical replicate processing in Maastricht)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection	Newly diagnosed, untreated pediatric UC/CD meeting Porto criteria vs healthy Dutch children; all IBD cases had active disease. Cross-hospital/platform analyses reprocessed stored DNA isolates from subsets of the same cohort.	n=149 original cohort: 91 IBD (53 CD, 38 UC), 58 healthy controls; age ≤18 y; exact age summary NR. Replicates: A2 n=116; M2 n=30.	Stool	IS-pro 16S-23S rRNA interspace profiling; ABI PRISM 3130 (A1) and ABI 3500 (A2/M2).	Taxonomic/community-profile features derived from IS fragment length and abundance.	Whole IS-pro profiles evaluated with seven supervised classifiers; random forest was the most robust to simulated and observed technical perturbations.	A1 10-fold CV: IBD vs healthy AUC 0.85-0.92 across classifiers; UC vs healthy maximal AUC=0.96; CD vs healthy maximal AUC=0.87.	Not evaluated in an independent patient cohort. Technical reproducibility only: the same stored DNA/subset samples were reprocessed across platform/hospital settings; RF cross-hospital accuracy was 1.00 on the 28 paired samples used for comparison.
35	Vermeer, 2026	40971415	Netherlands (Amsterdam Rapid cohort; LifeLines/Dutch Microbiome Project)	GI/hepatobiliary: Inflammatory bowel disease	Disease detection; Differential diagnosis; Phenotyping/stratification;	Therapy-naive children referred with suspected IBD/DGBI: de novo IBD vs symptomatic children in whom IBD was	n=534 baseline: 103 IBD (70 CD, 27 UC, 6 IBD-U), 75 CGI, 356	Stool	Shotgun metagenomics; Illumina NovaSeq 6000 S4, PE150; MetaPhlAn4 taxonomy and	Taxonomic species and functional pathway features.	Coda4Microbiome elastic-net generalized linear models using species/pathway ratios (maximum	Repeated random holdout validation: 75% training/25% testing, repeated for 50	No independent/external validation; all reported diagnostic and

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
					Prognosis/treatment-response prediction	excluded (CGI); age/sex-matched healthy LifeLines controls. Baseline microbiome also evaluated for CD vs UC and 3-month induction-treatment response.	healthy controls. Median age [IQR]: IBD 15 [13-16] y; CGI 14 [11.5-16]; healthy 13 [11-15]. Response analysis: 30 responders/37 nonresponders; nutritional subgroup 17/23.		HUMAnN functional profiling.		15 features); clinical-only and combined models; SVM/LASSO/RF sensitivity analyses.	permutations; 10-fold CV within training (5-fold when group n≤10). Coda4Microbiome species models: IBD vs healthy AUC=0.96±0.02; IBD vs symptomatic CGI=0.71±0.07 (RF sensitivity analysis 0.81±0.06); CD vs UC=0.74±0.11. For IBD vs CGI, the clinical model (age/sex/FCP/CRP) achieved AUC=0.90±0.05 and the combined model 0.80±0.06. Three-month response prediction was limited: combined AUC=0.63±0.09 overall and 0.67±0.13 for nutritional induction therapy.	treatment-response performance came from repeated splits of the study cohorts.
36	Luo, 2025	41430556	China (reported recruitment at a tertiary children's hospital in Beijing)	GI/hepatobiliary: Inflammatory bowel disease	Differential diagnosis	Children <18 y with clinically/endoscopically/histopathologically confirmed IBD vs non-IBD gastrointestinal controls (including acute self-limited gastroenteritis/other GI illness; healthy controls when available). The ML section additionally incorporated a Kaggle gastrointestinal dataset.	Prospective clinical cohort n=50: 30 IBD (18 CD, 12 UC), 20 controls; age 13.9±2.4 vs 14.3±2.1 y. Exact ML modeling denominator is Unclear because external Kaggle records were added and class prevalence no longer matches the 50-child cohort.	Stool	16S rRNA V3-V4 amplicon sequencing (341F/806R); Illumina MiSeq 2x250; QIIME2.	Taxonomic/community and co-occurrence-network features; clinical-microbiome composite models.	Random forest, logistic regression, XGBoost, and soft-voting classifier with SHAP; stable features included Faecalibacterium, Escherichia-Shigella, Veillonella, fecal calprotectin, and CRP/ESR.	Internal modeling reported 80: 20 stratified split with SMOTE in training. Table 5 test accuracy: RF 0.96, LR 0.85, XGBoost 0.98, voting 0.97. Repeated-CV XGBoost ROC-AUC: clinical 0.94 (0.91-0.97), microbiome 0.97 (0.95-0.98), combined 0.99 (0.98-0.99). Modeling denominator/source linkage is Unclear.	No independent clinical validation. The paper's prospective 50-child cohort and Kaggle-augmented ML dataset cannot be cleanly reconciled from the source report.
37	Papa, 2012	22768065	United States (Boston, Massachusetts)	GI/hepatobiliary: Inflammatory bowel disease	Differential diagnosis; Phenotyping/stratification	Children/young adults treated at Boston Children's gastroenterology program: IBD vs symptomatic non-inflammatory GI controls (e.g., abdominal pain, constipation, reflux, poor weight gain, diarrhea/blood in stool). Within IBD, CD vs UC and active vs remission were evaluated.	Initial cohort n=91: 23 CD, 43 UC, 1 IBD-U, 24 symptomatic non-inflammatory GI controls. Reported age values: CD 14.13±3.84 y (range 3–20), UC 13.7±4.25 (4–24), controls 9.08±4.3 (3–17).	Stool	16S rRNA V3-V5 amplicon sequencing; 454 FLX Titanium pyrosequencing.	Taxonomic/community-profile features.	SLiME (synthetic learning in microbial ecology) supervised classification/random-forest framework using microbial taxa.	SLiME random-forest classification using taxonomic relative-abundance profiles; no feature selection or dimensionality reduction in the classification task. Median AUC from three repetitions of 10-fold cross-validation: IBD vs symptomatic controls 0.83 (example threshold sensitivity	Independent blind validation cohort: n=68 according to Abstract, Results, and Methods (Figure S12 caption reports 77). IBD vs symptomatic controls AUC=0.84. Example ROC operating points included 92%

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
												80.3%, specificity 69.7%); active IBD vs controls 0.91; CD vs UC 0.76; active vs remission 0.72.	sensitivity/58.5% specificity and 94.5%/46.1% at different thresholds. Independent patient validation, but not clearly an external-center cohort.
38	Monleon-Getino, 2023	38034829	Spain (Barcelona)	GI/hepatobiliary: Inflammatory bowel disease (Crohn disease/ulcerative colitis)	Disease detection	Pediatric CD/UC diagnosed by Porto Criteria vs age/sex-matched children without known digestive disease. Baseline samples were collected before treatment; IBD participants were followed at 3 and 6 months.	Initially n=36 (12 CD, 12 UC, 12 healthy). Final reporting is internally inconsistent: abstract states n=30 but lists 13 CD+11 UC+8 healthy (=32); Results also gives 24 IBD+8 healthy. Inclusion age 6-18 y; median [range]: CD 11 [8-15], UC 13 [9-15], healthy 12 [10-16].	Stool; saliva	Shotgun DNA metagenomic sequencing; Nextera DNA Flex libraries; Illumina NovaSeq 6000 PE150; GAIA taxonomic profiling.	Taxonomic species-level features.	Three preselected 10-species fecal panels for baseline CD vs healthy controls; SVM, kernel discriminant analysis, and LDA.	Internal 80: 20 train/test split. Best reported scenario: Set 1 LDA training accuracy 1.00 (95% CI 0.85-1.00), test accuracy 1.00 (0.84-1.00), sensitivity/specificity 1.00/1.00. Other set/algorithm test accuracies ranged 0.59-0.95. Feature-selection/test-set independence is Unclear.	No independent/external validation.
39	Hollister, 2019	31005411	United States (Houston, Texas)	GI/hepatobiliary: Irritable bowel syndrome	Disease detection	Preadolescent children aged 7-12 y with Rome III IBS vs healthy controls, with detailed pain phenotyping.	n=45: 23 IBS, 22 healthy controls; age 7-12 y.	Stool	Whole-genome shotgun metagenomics (Illumina HiSeq 2000, 100-bp paired-end; MetaPhlAn2/HUMAnN2) plus untargeted fecal metabolomics (LC-MS/MS and GC-MS).	Multi-omics: taxonomic species, functional pathways, and fecal metabolites.	27 candidate multi-omics features were reduced to a 10-feature random-forest classifier; alternative classifiers were evaluated.	Fivefold cross-validation: IBS vs healthy AUC=0.93; accuracy=0.80. RF showed the best calibration among compared classifiers.	No independent/external validation.
40	Shankar, 2015	26653757	United States	GI/hepatobiliary: Irritable bowel syndrome, diarrhea-predominant	Disease detection	Age-matched US children/adolescents with diarrhea-predominant IBS vs healthy controls.	n=44 discovery: 22 IBS-D (average age 13.2 y), 22 healthy controls (12.6 y), both sexes. Independent extension: 4 newly recruited IBS-D adolescents only.	Stool	Phylogenetic microbiota microarray profiling plus fecal proton (1H) NMR metabolomics.	Multi-omics: microbial genus/phylotype features and fecal metabolites.	Separate microbiota- and metabolite-based PLS-DA classifiers combined using a Bayesian integration approach.	Reported cross-validation: genus-based PLS-DA accuracy=79.5% (sensitivity 72.7%, specificity 86.4%, AUC=0.88); metabolite-based model accuracy=81.8% (77.3%, 86.4%, AUC=0.87); Bayesian-integrated genus+metabolite model accuracy=84.1% (81.8%, 86.4%), AUC=0.93. A supplementary phylotype+NMR spectral-bin	Limited independent positive-case validation: the integrated model correctly classified all 4 newly recruited IBS-D adolescents. No independent healthy controls were included; external specificity, two-class accuracy, and AUC could not be estimated.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
												integrated model achieved accuracy=90.9% (40/44) and AUC=0.91.	
41	Saulnier, 2011	21741921	United States (Houston, Texas)	GI/hepatobiliary: Irritable bowel syndrome	Disease detection; Phenotyping/stratification	School-age children (7-12 y) meeting pediatric Rome III IBS criteria vs healthy children; IBS subtypes assigned from stool-form diaries.	n=44 children (22 IBS, 22 healthy). Seventy-one stool specimens were analyzed overall (29 healthy and 42 IBS in several V1–V3 analyses); some subjects contributed serial specimens.	Stool (repeated sampling)	16S rRNA amplicon profiling by deep 454/Roche pyrosequencing (primarily V1-V3; some V3-V5) and PhyloChip G2 microarray; qPCR confirmation for selected taxa.	Taxonomic/community features.	randomForest with Boruta feature selection; separate classification of IBS vs healthy and IBS-C vs IBS-U.	IBS vs healthy classification success >96%. Among IBS, IBS-C vs IBS-U classification reached up to 98.5% accuracy; no AUC/sensitivity/specificity reported.	No independent/external validation.
42	Xu, 2024	38127131	Public datasets; prospective validation (Tianjin, China)	GI/hepatobiliary: Biliary atresia	Differential diagnosis; Prognosis/treatment-response prediction	Discovery reanalysis: preoperative BA vs other cholestatic disease controls (DC) and healthy controls. Independent local cohort: infants with cholestasis undergoing surgery, BA/DC defined by intraoperative cholangiography, plus age-matched non-hepatobiliary controls; postoperative follow-up assessed jaundice clearance, cholangitis, native-liver survival.	Discovery: 118 fecal samples (43 BA, 33 DC, 42 healthy). Local cohort n=23 (11 BA, 6 DC, 6 healthy), cholestatic patients age ≤4 months; exact age distribution NR in the source report. Follow-up monthly for 6 months, then every 3-12 months to Dec 2022.	Stool; serum	16S rRNA V3-V4 amplicon sequencing/reanalysis; local fecal 16S sequencing; serum butyrate by liquid chromatography (C18 column; instrument details in Methods).	Taxonomic abundance ratios; metabolite (serum butyrate).	Random-forest screening followed by ROC analysis of genus abundance ratios; exploratory associations of microbiota/serum butyrate with postoperative outcomes.	Discovery BA vs DC: Streptococcus/Bacteroides AUC=0.9035 (95% CI 0.8347-0.9722); Streptococcus/Ruminococcus 0.8957; Streptococcus/Staphylococcus 0.7956; Streptococcus/Eggerthella 0.6258.	Independent local clinical cohort supported Streptococcus/Eggerthella AUC=0.8333 (95% CI 0.6340-1.000; P=0.027). This was candidate-ratio validation, not a locked multivariable classifier. Baseline serum butyrate correlations with jaundice clearance/cholangitis/liver fibrosis were not statistically significant; no prognostic AUC/model.
43	Lydon, 2026	41974724	United States (8 hospitals)	Respiratory: Lower respiratory tract infection	Differential diagnosis	Prospectively enrolled mechanically ventilated children with clinically adjudicated true LRTI, incidental pathogen carriage (IPC), or noninfectious respiratory failure (CTRL); intended to distinguish infection from carriage/noninfectious illness.	n=326 independent children: LRTI 207, IPC 70, CTRL 49. Median age [IQR]: 0.6 [0.2-2.1], 1.6 [0.9-6.4], and 9.5 [1.3-14.4] y, respectively. Same multicenter parent cohort as Mick et al. 2023 (PMID	Tracheal aspirate	Host-microbe metatranscriptomic RNA-seq; 150-bp paired-end Illumina NovaSeq; lower-airway FABP4 protein measured in a subset.	Multi-omics/host-microbe composite: host transcripts plus microbiome alpha diversity (Shannon); FABP4 protein as a pragmatic host biomarker.	Logistic models for single features/FABP4+Shannon; LASSO logistic multi-gene host model combined with Shannon diversity.	Stratified 5-fold CV with out-of-fold predictions: FABP4+Shannon AUC=0.87 (95% CI 0.83-0.91); multi-gene+Shannon AUC=0.89 (0.85-0.92). FABP4 protein subset n=134: AUC=0.88 (0.82-0.93).	Not evaluated in a separate external cohort; performance was internal cross-validation within the multicenter parent cohort.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							37009900; No. 44).						
44	Mick, 2023	37009900	United States (8 hospitals)	Respiratory: Lower respiratory tract infection	Differential diagnosis	Prospective mechanically ventilated pediatric acute respiratory failure cohort. Classifier development used clinically adjudicated Definite LRTI vs No Evidence of LRTI/noninfectious respiratory failure; Suspected/Indeterminate cases were analyzed separately without a definitive reference label.	n=261 with high-quality host/microbial data: Definite LRTI 117, No Evidence 50, Suspected 57, Indeterminate 37. Median age [IQR]: 0.5 [0.2-1.8], 6.5 [1.5-12.9], 1.7 [0.5-6.0], 1.5 [0.6-10.8] y. Same parent cohort as Lydon et al. 2026 (PMID 41974724; No. 43).	Tracheal aspirate	Metagenomic RNA-seq of tracheal aspirate; 150-bp paired-end Illumina sequencing; host transcriptome and respiratory microbiota/pathogen profiling.	Multi-omics/host-microbe composite: host gene-expression probability, respiratory-virus abundance, and pathogenic bacterial/fungal dominance.	Host LRTI classifier: LASSO gene selection within training folds + random forest; integrated classifier added microbial/pathogen features.	Fivefold cross-validation: host classifier median AUC=0.967 (range 0.953-0.996); integrated classifier median AUC=0.986 (0.953-1.000).	No independent/external validation. Application to 94 Suspected/Indeterminate patients (52% called LRTI; likely pathogen nominated in 98% of predicted LRTI) is not validation because reference status was uncertain.
45	Li, 2019	31494301	China (Qingdao, Shandong)	Respiratory: Recurrent respiratory tract infection	Disease detection	Children meeting 2007 Chinese Pediatric Society RRTI criteria vs healthy volunteers; exclusions included recent antibiotics/probiotics and major chronic, GI, congenital, or immunodeficiency conditions.	n=49: 26 RRTI, 23 healthy controls. Age was recorded but exact group age statistics are NR in the source report.	Stool	16S rRNA V3-V4 amplicon sequencing (338F/806R); Illumina MiSeq paired-end 2x250.	Taxonomic/community features.	Random forest identified a 17-genus RRTI-associated signature; ROC analysis was reported for individual Enterococcus, Eubacterium, and Bifidobacterium.	Same-cohort ROC: Enterococcus AUC=0.860; Eubacterium=0.820; Bifidobacterium=0.689. No validated multivariable RF performance was reported.	No independent/external validation.
46	Duan, 2026	42210054	Public SRA secondary analysis; pediatric source geography NR in the source report	Respiratory: Respiratory syncytial virus infection	Disease detection; Phenotyping/stratification	Secondary analysis of public pediatric RSV gut-microbiome datasets with control, infected, and recovery states; recovery samples were collected 15-150 days post-discharge. Mouse datasets were analyzed separately for cross-species comparison.	319 total samples: 154 pediatric and 165 mouse. Pediatric group denominators, unique-child count, age distribution, and source geography are NR in the source report.	Stool (secondary analysis)	Public 16S rRNA amplicon datasets with heterogeneous original single/paired-end sequencing; uniformly reprocessed using QIIME2/DADA2 and SILVA v138.99; PICRUSt2 functional prediction.	Taxonomic family/genus features; predicted functional pathways.	13 machine-learning algorithms at family/genus levels; top-30 importance features per model; cross-species shared signatures.	Random sample split 70% training/30% validation. Pediatric genus-level validation AUC: CatBoost=0.952; random forest=0.937. Models also separated infected, recovery, and control states; task-specific binary values are in supplementary material.	No independent external cohort. The 30% random validation set came from the same integrated public datasets; independence at participant/project level was not established.
47	Henares, 2021	34699345	Spain (Barcelona)	Respiratory: Acute respiratory infection/respiratory health	Disease detection; Differential diagnosis	Prospective case-control study: invasive pneumococcal disease (IPD) vs symptomatic mild viral URTI controls and healthy/asymptomatic controls. IPD was confirmed by S. pneumoniae detection in normally sterile samples.	n=140: 27 IPD, 48 viral URTI, 65 healthy/asymptomatic. Median age [IQR]: IPD 33 [19.0-49.5] months; URTI 24.5	Nasopharyngeal aspirate	16S rRNA V3-V4 amplicon sequencing, Illumina MiSeq; pneumococcal qPCR/quantification/serotyping; multiplex respiratory-virus PCR.	Taxonomic/community plus microbiological composite features.	Three random-forest models using all microbiological variables: IPD vs viral URTI; IPD vs healthy; IPD vs all controls.	Apparent/internal model performance: IPD vs viral URTI AUC=0.90 (95% CI 0.83-0.97); IPD vs healthy AUC=0.86 (0.78-0.94); IPD vs all controls AUC=0.86 (0.78-0.93).	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							[14.7-45.0]; healthy 31 [19.0-43.0].						
48	Wang, 2023	37132354	China (Beijing)	Respiratory: Mycoplasma pneumoniae pneumonia	Differential diagnosis; Phenotyping/stratification	Children with community-acquired pneumonia undergoing bronchoscopy: MPP vs bacterial/viral pneumonia controls. MPP classification used MP RNA, MP-specific antibody, or clinical diagnosis; severity was assessed within MPP.	n=187: 158 MPP (99 RNA-positive, 51 antibody-positive, 8 clinically diagnosed), 29 other bacterial/viral pneumonia. MPP median age 7.6 y [IQR 6.0-9.6], range 1-16; controls 4.6 [1.0-10.0], range 0.2-13. MPP: 101 mild, 57 severe.	Bronchoalveolar lavage fluid	16S rRNA V3-V4 amplicon sequencing (341F/806R); Illumina NovaSeq.	Taxonomic/community features; primary diagnostic marker was Mycoplasma relative abundance.	Single-genus Mycoplasma abundance threshold 0.126% for MPP vs other CAP; severity analyses compared diversity and Mycoplasma abundance in mild vs severe MPP.	Same-cohort ROC: MPP vs other CAP AUC=0.994; sensitivity 97.5% (154/158), specificity 96.6% (28/29). Severe MPP had lower alpha diversity and higher Mycoplasma abundance; no severity classifier performance was reported.	No independent/external validation.
49	Gao, 2026	42365232	China (Hohhot, Inner Mongolia)	Respiratory: Mycoplasma pneumoniae pneumonia	Differential diagnosis; Phenotyping/stratification	Pediatric respiratory-disease patients: clinically diagnosed MPP vs MPP-negative children with other respiratory diseases (e.g., bronchitis, lobar/EBV pneumonia, asthma). MPP severity and respiratory-site strata were also analyzed.	n=688: 401 MPP, 287 non-MPP. Median age 9.1 vs 6.7 y; ranges 1.2-17.0 vs 0.1-17.0. MPP: 280 mild/121 severe; non-MPP 231/56. MPP upper/lower respiratory tract samples: 233/168.	Pharyngeal swab; sputum; bronchoalveolar lavage fluid	16S rRNA V3-V4 amplicon sequencing (338F/806R); Illumina NovaSeq 6000 SP, paired-end.	Taxonomic/community features.	Random forest ranked 26 differential genera; top 15 were examined further. Diagnostic discrimination was based on single-genus Mycoplasmoides relative abundance (cutoff 0.0466%), rather than a multivariable RF classifier. Separate exploratory analyses assessed severity and URT-vs-LRT patterns.	Same-cohort apparent ROC: Mycoplasmoides AUC=0.983, sensitivity=95.3%, specificity=98.3%, cutoff=0.0466%. No train/test split, cross-validation, bootstrap validation, or held-out evaluation was reported. Severe MPP showed lower Chao1/Shannon diversity and higher Mycoplasmoides abundance; microbiome–clinical correlations did not remain significant after Benjamini-Hochberg correction. For anatomical-site discrimination, Mycoplasmoides distinguished LRT from URT samples with AUC=0.928, sensitivity=88.7%, specificity=96.6% (exploratory same-cohort analysis).	Not evaluated; no independent/external cohort. The diagnostic findings were exploratory and explicitly require external validation.
50	Moser, 2026	42589087	Austria (Graz)	Respiratory: Adenoid hypertrophy	Disease detection	Children aged 2-10 y with symptomatic pathologic adenoid hypertrophy confirmed intraoperatively at planned adenoidectomy (≥3 typical symptoms) vs age-matched children with non-hypertrophic adenoids	Enrolled n=60 (30 AH, 30 controls); final microbiome analysis n=56 after 4 low-depth AH samples were	Adenoidal mucus / nasopharyngeal secretion collected endoscopically during surgery	16S rRNA V4 amplicon sequencing (515FB/806RB); Illumina MiSeq; QIIME2/DADA2 with SILVA 138.	Taxonomic/community features (genus-level abundance/diversity).	Random forest classifier (1, 000 trees; mtry=7) using genus-level profiles; importance features included Escherichia-Shigella, Comamonadaceae, Cutibacterium,	Internal out-of-bag evaluation only: OOB error=5.36%; all 30 controls were correctly classified and 3/26 AH samples were misclassified. No ROC-AUC was	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						undergoing unrelated noninfectious elective surgery; recent URTI/antibiotic/corticosteroid exposure and chronic respiratory disease were excluded.	excluded: 26 AH, 30 controls. Age mean ± SD: AH 5.4 ± 2.1 y; controls 5.0 ± 2.1 y.				Paludibacteraceae, Nocardia, and Fusobacterium.	reported.	
51	Liu, 2023	38090578	China; discovery: East China (Zhejiang/Jiangsu/Shanghai); validation geography internally inconsistent in the article (Results: Gansu/Tibet/Guangdong; Discussion: Sichuan/Guangdong/Yunnan/Tibet)	Respiratory: Adenoid hypertrophy	Disease detection	Children with AH confirmed by nasendoscopy with >75% obstruction vs healthy controls without rhinitis, snoring, or adenotonsillar hypertrophy. Recent antibiotics/probiotics/prebiotics, recent surgery, and major inflammatory/GI disorders were excluded.	Total n=219 (119 AH, 100 controls). Discovery n=130 (77 AH, 53 controls); geographic validation n=89 (42 AH, 47 controls). Age mean ± SD: AH 4.59 ± 1.42 y; controls 4.86 ± 1.59 y; age-stratified analyses used 2-3, 4-5, and 6-8 y groups.	Stool	16S rRNA V3-V4 amplicon sequencing (338F/806R); Illumina MiSeq; OTU-based analysis; PICRUSt2 functional prediction.	Taxonomic OTU/community features plus predicted microbial functions.	Random forest selected an 8-OTU microbial marker panel by 5-fold cross-validation; probability-of-disease (POD) index used for discrimination.	Discovery cohort: AUC=0.9851 (95% CI 0.9654-1.000; P<0.0001), AH n=77 vs controls n=53.	Geographically different validation cohort: same 8-marker panel/POD framework AUC=0.9782 (95% CI 0.9476-1.000; P<0.0001), AH n=42 vs controls n=47. Whether model parameters/threshold were fully locked before validation was not clearly reported.
52	Huang, 2022	36105146	China (Beijing)	Respiratory: Obstructive sleep apnea/hypopnea	Disease detection; Phenotyping/stratification	OSA children aged 6-12 y with overnight PSG-confirmed OSA (OAHI ≥1 event/h), adenotonsillar hypertrophy, and mouth breathing vs low-risk controls recruited from hospital-staff families with PSQ-SRBD score <3; controls were not PSG-confirmed. Eight OSA children were reassessed 1 y after adenotonsillectomy.	Cross-sectional n=58: 36 OSA (age 7.47 ± 2.24 y) and 22 controls (7.55 ± 2.48 y). Longitudinal postoperative subset n=8; pretreatment age 8.8 ± 2.0 y.	Saliva	16S rRNA V3-V4 amplicon sequencing; Illumina NovaSeq, paired-end 2 x 250 bp; QIIME2/DADA2; Human Oral Microbiome Database taxonomy.	Taxonomic/community features (genus/species-level salivary microbiota).	Nested cross-validated random forest using 10 genera: Lactobacillus, Escherichia, Bifidobacterium, Capnocytophaga, Bacteroidetes_[G-7], Parvimonas, Bacteroides, Klebsiella, Lautropia, and Prevotella. Prevotella sp_HMT_396 was also examined across OSA severity and longitudinally.	OSA vs controls: nested-CV RF AUC=0.94. Prevotella sp_HMT_396 abundance increased across OSA severity (P=0.014) and was lower 1 y after adenotonsillectomy in the 8 paired patients (P=0.003); no severity classifier performance was reported.	No independent/external validation; the 1-y postoperative subset was longitudinal follow-up of the same cohort, not external validation.
53	Wu, 2022	35580999	China (Ningbo, Zhejiang)	Respiratory: Obstructive sleep apnea/hypopnea	Disease detection	Children aged 4-6 y presenting with sleep snoring and diagnosed with OSAHS according to Chinese guidelines after PSG vs healthy age-matched children; all subjects underwent PSG. Children with obesity, allergy/asthma, major comorbidities, or recent antibiotics/probiotics were excluded.	n=88: 43 OSAHS, 45 controls. Age mean ± SD: 5.08 ± 0.53 vs 4.91 ± 0.56 y.	Stool	16S rRNA V4 amplicon sequencing (515F/806R); Illumina MiSeq; OTU clustering at 97%; PICRUSt functional prediction.	Taxonomic/community features plus predicted functional pathways.	Three-genus combined microbiota model using Faecalibacterium, unclassified Ruminococcaceae, and Roseburia; individual and combined ROC analyses were reported.	Same-cohort ROC: combined model AUC=0.794 (SE 0.048; 95% CI 0.700-0.888), sensitivity 79.1%, specificity 80.0%. Individual AUCs: Faecalibacterium 0.689; unclassified Ruminococcaceae 0.641; Roseburia 0.539. No resampling/holdout validation was reported.	No independent/external validation.
54	Novau-Ferre, 2025	40184961	Spain (Tarragona)	Neuro/psych: ASD/ADHD	Differential diagnosis	Children aged 5-14 y with DSM-5 ADHD vs DSM-5 ASD enrolled in a 12-week randomized double-blind probiotic/placebo	Randomized n=80 (39 ADHD, 41 ASD); baseline microbiota	Stool	16S rRNA V4 amplicon sequencing; Ion Torrent Personal Genome Machine;	Taxonomic features plus predicted KEGG functions.	Elastic-net penalized logistic regression on 22 discriminatory taxa (21 genera plus class	Baseline ADHD vs ASD: AUC=0.939 (95% CI 0.937-0.941), sensitivity=0.872 (0.866-0.879),	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						trial. The diagnostic comparison used baseline microbiota; no healthy-control group was included. Post-intervention microbiota changes were treatment effects, not baseline treatment-response prediction.	available for n=77 (exact ADHD/ASD analytic counts NR). Baseline age median [IQR]: ADHD 10.00 [7.00, 12.00] y; ASD 10.00 [7.50, 11.00] y.		DADA2/QIIME2 with SILVA 138; PICRUSt2 functional prediction.		Clostridia), adjusted for age, sex, BMI z-score, and medication; signatures retained taxa consistently selected across 100 iterations.	specificity=0.923 (0.917-0.928).	
55	Cai, 2025	40045873	China (Shandong main cohort); prior published ASD/ADHD datasets used for validation	Neuro/psych: ASD/ADHD	Disease detection	Children aged 2-11 y with clinician-confirmed DSM-5 ASD, ADHD, or both ASD+ADHD vs unrelated age- and sex-matched healthy controls. Separate previously published ASD and ADHD datasets were used as validation sets; independence/participant overlap with the current cohort is not fully documented in the source report.	Main cohort n=284: 113 ASD, 43 ADHD, 8 ASD+ADHD, 120 healthy controls; age 2-11 y. Exact validation-set sample sizes are NR in the source report.	Stool	16S rRNA sequencing with diagnosis-aligned amplicon protocols: ASD/healthy V4; ADHD/ASD+ADHD V3-V4; Illumina HiSeq4000, paired-end 2 x 150 bp; QIIME; PICRUSt/KEGG and gut-brain-module functional profiling.	Taxonomic/community features plus predicted functional KO/pathway and gut-brain-module profiles.	Random-forest classifiers (1, 000 trees) using all taxa, fully annotated taxa, or 16 selected genera; 10-fold cross-validation. Selected microbial functions were also modeled.	Main cohort, patients vs controls: 10-fold CV AUC all taxa=0.986 (95% CI 0.97-1.00), fully annotated taxa=0.986 (0.97-1.00), selected 16 genera=0.958 (0.94-0.98); reported model accuracy range 0.92-0.97. The mixed V4 vs V3-V4 sequencing scheme is aligned with diagnostic groups and is a potential technical confounder.	Previously published validation datasets: selected-genus models reported AUCs 0.69-0.74 (abstract; subgroup values shown graphically).
56	Tian, 2026	41781872	China (Hebei Province)	Neuro/psych: Migraine	Disease detection; Phenotyping/stratification	Treatment-naive children aged 5-14 y meeting ICHD-3 migraine criteria with symptoms for ≥1 month vs healthy controls from the same province; migraine patients were also stratified by attack frequency (episodic, high-episodic, chronic).	n=60: 30 migraine, 30 healthy controls; 60 fecal samples. Age mean ± SD: 10.47 ± 2.53 vs 9.33 ± 2.12 y.	Stool (same fecal samples used for metagenomics and untargeted metabolomics)	Shotgun metagenomic sequencing; Illumina NovaSeq X Series 25B, 300-cycle; fecal untargeted LC-MS using Thermo Fisher UHPLC-Q Exactive HF-X.	Multi-omics: taxonomic species, microbial functional genes/pathways, and fecal metabolites.	ROC analyses of microbial functional pathways; lipopolysaccharide (LPS) biosynthesis was the strongest reported pathway marker. A random-forest model integrating differential microbial species and metabolites was described in Methods, but quantitative performance for that integrated model was NR.	Same-cohort ROC: LPS biosynthesis AUC=0.75 (95% CI 0.63-0.87); other pathway AUCs were lower (approximately 0.54-0.68). Attack-frequency differences were exploratory and none remained significant after FDR correction. No holdout/CV scheme was reported for the ROC analyses.	No independent/external validation.
57	Liu, 2024	38177986	United States (GMrepo/American Gut Project microbiome cohort); China (Qilu Hospital, Shandong, plasma metabolite cohort)	Neuro/psych: Migraine	Disease detection; Phenotyping/stratification	Two separate pediatric cohorts: secondary gut-microbiome data for migraine vs healthy children from GMrepo/American Gut Project; and a Qilu Hospital plasma cohort of ICHD-3 migraine vs healthy children. Plasma analyses also compared migraine with/without aura and ictal/interictal states.	Microbiome cohort n=75: 33 migraine (age 7.48 ± 1.95 y), 42 healthy (7.57 ± 2.23 y). Plasma metabolite cohort n=171: 51 migraine (10.75 ± 2.20 y), 120 healthy (9.91 ± 5.33 y).	Stool (secondary microbiome data); plasma	Secondary analysis of GMrepo/American Gut Project gut microbiota data (original 16S region/platform NR in source report); plasma TRP, 5-HT, KYN, KYNA, and QUIN measured by ELISA.	Taxonomic gut microbial genera plus plasma tryptophan-pathway metabolites.	Microbial models: 7-genus ridge-regression panel; 28-genus LEfSe panel; 19-genus random-forest panel. Plasma models: KYNA/QUIN ratio and combined KYNA+QUIN+5-HT indicator. Microbiome and plasma models were derived in different participant cohorts.	Same-cohort microbial performance: ridge 7-genus AUC=0.973, sensitivity 97.6%, specificity 90.9%; LEfSe 28-genus AUC=0.988, sensitivity 100%, specificity 93.9%; RF 19-genus AUC=0.994, sensitivity 97.6%, specificity 97.0%. Plasma Results: KYNA/QUIN AUC=0.871 (95%	No independent/external validation. The separate U.S. microbiome and Chinese plasma cohorts validate different biomarker modalities and do not constitute external validation of the same model.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
												CI 0.818-0.924), sensitivity 86.7%, specificity 78.4%; KYNA+QUIN+5-HT AUC=0.897 (0.851-0.942), sensitivity 86.3%, specificity 83.3%. Abstract instead assigns 86.3%/83.3% to KYNA/QUIN (internal inconsistency).	
58	Zeng, 2025	39855566	China (Luzhou, Sichuan)	Neuro/psych: Major depressive disorder	Disease detection	Treatment-naive adolescents with first-episode DSM-5 MDD (MINI-assessed; CDI ≥19) recruited from psychiatric inpatient/outpatient units vs adolescents without prior psychiatric or major physical illness recruited from nearby schools; both groups age 12-17 y.	n=74 (37 MDD, 37 HC). Age, mean ± SD: MDD 14.51 ± 1.66 y; HC 16.08 ± 1.09 y.	Saliva	Bacterial 16S rRNA amplicon sequencing (338F/806R); sequencing instrument NR in the source report; detailed saliva/model methods NR in the source report.	Taxonomic (ASV/genus-level) and community/diversity.	Random forest ranking; top 10 predictive oral microbial features used as an ASV set for ROC analysis; exact taxa/model-validation workflow is partly NR in the source report.	Internal/same-cohort analysis: AUC=0.78 (95% CI 0.67-0.88). Exact resampling scheme and whether feature selection was nested are NR in the source report.	No independent/external validation.
59	Xi, 2021	33738808	NR in the source report (participant recruitment location NR in the source report)	Neuro/psych: Tic disorders	Disease detection	Children with tic disorders (PTD, CTD, or Tourette syndrome; 35 treatment-naive and 14 medicated) vs matched healthy controls. DRA exposure was examined secondarily within TD.	n=99 (49 TD, 50 HC). TD age mean 8.84 ± 2.35 y, range 4-14; HC age NR in the source report. TD subtypes: PTD 17, CTD 9, TS 23.	Stool	Shotgun metagenomic sequencing; sequencing instrument/library details NR in the source report in the source report.	Taxonomic species-level profiles and functional/metabolic pathways.	Random forest classifier using bacterial species abundance; top-10 species importance reported, but no fixed final panel stated.	Internal only: 65% randomly selected samples used as training set; five-fold CV reported; AUC=0.884 (95% CI 0.778-0.990) for 49 TD vs 50 HC. Use of the remaining 35% and exact resampling/nesting are Unclear from source report.	No independent/external validation.
60	Li, 2022	35157224	China (Shanghai)	Neuro/psych: Viral encephalitis/meningitis	Differential diagnosis	Children undergoing lumbar puncture: clinically diagnosed viral encephalitis and/or viral meningitis (VEVMC) vs frequency-matched children with other diseases (e.g., fever, headache, epilepsy, febrile convulsion); etiologically positive VEVM (VEVMP) analyzed secondarily.	512 recruited; 398 eligible; VEVMC n=68 (VEVMP PCR/RT-PCR positive n=43; etiologically negative n=25); matched other-disease controls n=68. Age eligibility 1 month-16 y.	Buccal mucosal oral swab	16S rRNA V3-V4 amplicon sequencing (341F/806R); Ion S5 XL; 600-bp single-end reads.	Taxonomic genus-level oral microbiome and predicted pathways.	Random forest: 18 genera for VEVMC vs controls; 8 genera for VEVMP vs controls.	Internal resampling: 10-fold CV (5 replicates), with bootstrap-like two-thirds training by sampling with replacement and remaining samples for validation. VEVMC: AUC=0.920 (95% CI 0.834-1.000), accuracy 90.7%, sensitivity 86.4%, specificity 95.2%, PPV 95.0%, NPV 87.0. VEVMP: AUC=0.945 (0.840-1.000), sensitivity 91.7%, specificity 95.5%.	No independent/external validation.
61	Li, 2025	41019533	China (Wenzhou, Zhejiang); external transfer dataset from a European pediatric cohort	Neuro/psych: Epilepsy	Differential diagnosis; Phenotyping/stratification	Pediatric common epilepsy (CEP) and refractory epilepsy (REP) vs symptomatic non-epilepsy (NEP) controls ultimately diagnosed with functional headaches; NEP	n=50: CEP 17, REP 23, NEP 10. Age mean ± SD: CEP 9.06 ± 3.21 y; REP 8.52 ± 2.57 y;	Stool; cerebrospinal fluid	Stool 16S rRNA V3-V4 amplicon sequencing (341F/805R), Illumina NovaSeq, PE250; CSF untargeted UHPLC-MS/MS, Vanquish	Taxonomic genera; predicted microbial functions; CSF metabolites; microbiota-metabolite	RF models: 8 genera (CEP vs NEP); 14 genera (REP vs NEP); selected CSF metabolites (REP vs NEP); integrated 10 genera + 4	LOOCV, plus 1, 000 permutations. CEP vs NEP: ROC-AUC 0.906, PR-AUC 0.942, accuracy 0.852. REP vs NEP: ROC-AUC 0.913,	External pediatric European 16S dataset: CEP-trained model applied to

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						lumbar puncture excluded CNS/neurologic involvement. CEP and REP represent disease phenotypes.	NEP 11.2 ± 3.58 y. CSF metabolomics subset: 16 REP and 8 NEP (n=24).		UHPLC-Q Exactive HF-X Orbitrap.	composite.	correlated CSF metabolites.	PR-AUC 0.965, accuracy 0.818. CSF metabolite model REP vs NEP: ROC-AUC 0.875, PR-AUC 0.943, accuracy 0.833. Integrated paired model: ROC-AUC 0.953, PR-AUC 0.976, accuracy 0.875. Differential-feature selection appears to precede LOOCV; nesting was not reported.	medication-sensitive epilepsy vs HC, AUC=0.838, accuracy=0.812; REP-trained model applied to medication-resistant epilepsy vs HC, AUC=0.796, accuracy=0.723. External n NR in the source report; only 7/8 and 9/14 original genera were present, so cross-platform feature handling should be interpreted cautiously.
62	Liu, 2023	36713168	China (Xi'an, Shaanxi)	Neuro/psych: Epilepsy	Disease detection; Phenotyping/stratification	Infants with newly diagnosed epilepsy (EP), epilepsy with diarrhea (ED), and ED after 2 weeks of Clostridium butyricum MIYAIRI 588 probiotic treatment (PT) vs healthy infants (CK). PT represents a post-treatment state, not an independent cohort.	Article states 20 infants with epilepsy plus 7 healthy controls, but reports CK n=7, EP n=13, ED n=13, PT n=13 (46 fecal samples/time points); exact unique-participant mapping is Unclear. Age, mean ± SD months: CK 4.9 ± 2.5; EP 4.6 ± 2.7; ED 4.2 ± 2.4; PT 4.5 ± 2.6.	Stool	16S rRNA V3-V4 amplicon sequencing (341F/805R); Illumina NovaSeq, PE250.	Taxonomic ASV-level/community features.	Random forest multiclass classifier; top 7 ASVs, including Escherichia/Shigella and Bifidobacterium as highest-importance markers.	Internal multiclass evaluation: all-ASV RF classifier accuracy=67.39%; selected top-7-ASV classifier accuracy=71.74%.	Not evaluated; no independent/external validation cohort.
63	Manghi, 2024	39528484	United States (SPARK-WGS family cohort across diverse US locations)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with professionally diagnosed ASD vs neurotypical biological siblings from SPARK families; family relatedness, sex, age, ancestry, and read depth were addressed in analyses.	SPARK-WGS n=7,812 participants from 2, 025 families; child classification subset: 2, 154 ASD (mean age 9 y [8.6, 9.1]) and 1, 646 NT siblings (mean age 8 y [7.8, 8.2]).	Saliva (secondary analysis of non-human reads from host whole-genome sequencing)	Deep whole-genome sequencing of saliva with secondary metagenomic analysis of non-human reads; MetaPhlAn 3/StrainPhlAn 3/HUMAnN 3.0; sequencing instrument NR in the source report.	Taxonomic species/strain-level and functional pathway/neuroactive-potential features.	Species-abundance random forest classifier; 108 species were independently associated with ASD/NT status (q<0.005), but these were not presented as a fixed 108-species diagnostic panel.	100 balanced datasets (500 ASD + 500 NT, max one child/family) with 10-fold, 10-times repeated CV; mean ROC-AUC=0.66. RF feature importance was calculated from training folds only.	Not evaluated: no independent/external classifier validation. External oral datasets were used for dataset-composition comparison, not validation of the ASD classifier.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
64	Hicks, 2018	30107083	United States (specific recruitment-site distribution NR in the source report)	Neuro/psych: Autism spectrum disorder	Disease detection; Differential diagnosis; Phenotyping/stratification	Children age 2-6 y: clinician-consensus DSM-5 ASD vs typically developing (TD) children and nonautistic developmental delay (DD); within ASD, GI-disturbance status was classified secondarily.	n=346: ASD 180, TD 106, DD 60. GI phenotype: 39 ASD with GI disturbance vs 141 without. Age 2-6 y.	Saliva (sublingual/parotid oral collection)	Salivary RNA shotgun metatranscriptomic sequencing; Illumina NextSeq 500; single-end 50-bp reads, target ~10 million reads/sample.	Taxonomic transcriptional activity and microbial functional pathways.	Multivariable logistic regression using microbial abundance ratios. ASD vs TD: 5 ratios/8 taxa; ASD vs DD: 3 ratios/5 taxa; ASD GI+ vs GI-: 3 ratios/5 taxa.	Internal 50/50 split. Training/CV -> naive holdout AUCs: ASD vs TD 0.795 (95% CI 0.711-0.872) -> 0.796; ASD vs DD 0.770 (0.643-0.867) -> 0.765; GI+ vs GI- 0.839 (0.759-0.958) -> 0.857. Holdout counts were approximately ASD 90, TD 53, DD 30; GI+ 20 vs GI- 71.	Not evaluated: no independent/external cohort; the 50% holdout is an internal random split.
65	Qiao, 2018	29371629	China (Shanghai)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with DSM-5 ASD, confirmed using ICD-10 criteria, vs unrelated age-/sex-matched healthy school controls; no prior ASD medical treatment other than rehabilitation.	n=59 children (32 ASD, 27 controls), yielding 111 specimens. Saliva: 32 ASD/27 controls; dental plaque: 26 ASD/26 controls. Age range 7-14 y; saliva age mean ± SD 10.02 ± 1.43 vs 10.19 ± 0.59 y.	Saliva; supragingival dental plaque	16S rRNA V3-V4 amplicon sequencing (338F/806R); Illumina MiSeq, PE250.	Taxonomic OTU-level/community signatures.	Random forest with 10-fold CV for marker-number selection; microbial index of autism (MIA). Saliva optimal combination: 20 OTUs (7 ASD-enriched + 13 control-enriched); dental plaque: 6 control-enriched OTUs.	Same-cohort MIA ROC after marker/combination selection: saliva AUC=0.963; cutoff -482.5; sensitivity 0.889; specificity 0.906. Exact dental-plaque AUC is NR in the source report. The final MIA ROC was not reported as an independent holdout estimate.	No independent/external validation.
66	Tang, 2025	39603332	Hong Kong SAR, China	Neuro/psych: Autism spectrum disorder	Disease detection	Preschool children with clinician-diagnosed DSM-5 ASD vs age-/sex-matched typically developing children screened negative on SRS-2; recruited from local preschools and rehabilitation centers.	64 recruited (32 ASD, 32 TD); 55 plaque samples had sufficient DNA and were sequenced/analyzed (25 ASD, 30 TD). Age mean ± SD: ASD 4.79 ± 0.83 y; TD 4.90 ± 0.85 y.	Pooled supragingival dental plaque	16S rRNA V3-V4 amplicon sequencing; Illumina NovaSeq PE250.	Taxonomic species-level oral microbiome.	11 discriminatory species (6 ASD-associated, 5 TD-associated) in a random forest model.	Internal split: 70% training; LOOCV used during model development; random testing set validation: accuracy 0.813 (95% CI 0.544-0.960), sensitivity 0.778, specificity 0.857, ROC-AUC 0.937 (95% CI 0.82-1.00).	No independent/external validation.
67	Shen, 2026	41953088	China (Suzhou, Jiangsu)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD from a special rehabilitation institution vs healthy kindergarten controls; ASD diagnoses had been made by professional psychologists/psychiatrists. Groups differed in age.	n=101 (54 ASD, 47 HC). Age mean ± SD: ASD 4.5 ± 1.4 y; HC 5.6 ± 0.5 y.	First-morning urine	Targeted chromatography/mass-spectrometry panel: HPLC-UV creatinine; LC-MS cortisol/cortisone; HPLC B vitamins; GC-MS SCFAs; amino acids by 3200 QTRAP MS/SRM. Exact instrument details for some assays are referenced to prior methods.	Metabolites related to the microbiota-gut-brain axis (indirect microbiome-related metabolic markers, not direct microbial measurements).	SVM combining 7 indicators: cortisol/cortisone ratio, creatinine, cortisol, taurine, histamine, homocysteine, methionine; age effects were addressed before biomarker analysis.	Internal analysis: ROC-AUC=0.943 (95% CI 0.855-0.997), sensitivity 92.6% (50/54), specificity 93.7% (44/47); average accuracy 88.9% over 100 cross-validations; permutation P<0.01. Upstream differential-feature selection preceded model evaluation; nesting	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
												was not reported.	
68	Flynn, 2026	42192173	United States (Arizona, Massachusetts, Tennessee, Texas)	Neuro/psych: Autism spectrum disorder	Disease detection	Children age 2-11 y with prior clinical ASD diagnosis (DSM-5; 51/52 additionally CARS-confirmed) vs healthy typically developing children recruited across four US sites.	Recruited n=99 (52 ASD, 47 TD), age mean ± SD 6.7 ± 2.8 vs 6.7 ± 3.0 y. Semiquantitative analysis: 50 ASD +47 TD; targeted quantitative analysis: 52 ASD +47 TD; five participants did not overlap between phases.	Urine	Discovery: semiquantitative untargeted LC-MS, Orbitrap Fusion Lumos Tribrid; targeted quantitative phase at a separate laboratory using HPLC/Agilent 6530 QTOF with standards.	Microbially derived urinary metabolites (phenylalanine-, tryptophan-, yeast/other-related).	MDM System score: number of selected MDMs above the maximum TD reference value; semiquantitative system used 17 MDMs. Additional FDA/neural-network/Naive Bayes analyses are detailed in supplement.	Semiquantitative same-cohort MDM System: sensitivity 90% (45/50 ASD), specificity 100% (47/47 TD). Targeted quantitative phase: sensitivity 78%, specificity 100%. The 100% specificity is partly inherent to defining elevation as above the maximum value in the same TD reference set. The two phases used partially overlapping participants and therefore do not constitute independent clinical validation.	Not evaluated: no independent validation cohort; authors state independent replication is needed/underway. Multi-site recruitment and separate analytical laboratory do not by themselves establish external clinical validation.
69	Xiong, 2016	27123458	China (Changsha, Hunan)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with DSM-IV ASD vs age-/sex-matched non-ASD controls; cases had no food-restriction history and controls excluded neurodevelopmental/psychiatric conditions including ADHD and tics.	n=124 (62 ASD, 62 controls), age 1.5-7 y; ASD 48 male/14 female and matched controls. Separate mechanistic/treatment analysis included 50 HPHPA-positive ASD children given oral vancomycin.	Urine	GC-MS metabolomics; Agilent 7890-5975C GC-MS with HP-5 capillary column.	Putatively gut-microbiota-derived aromatic metabolites.	Three-metabolite panel: HPHPA, 3HPA, 3HHA; prediction probability generated by the article-described "two-regression analysis.".	Same-cohort ROC: combined AUC=0.962, sensitivity 90.3%, specificity 98.4%; each individual metabolite had >96% specificity. No cross-validation/independent test set reported.	Not evaluated: no independent diagnostic validation. Vancomycin-associated metabolite decreases in 50 HPHPA-positive ASD children support putative microbial origin but are not classifier validation or prospective treatment-response prediction.
70	Liu, 2026	41748159	Multicohort: Hong Kong SAR in-house plus public pediatric cohorts from China, USA, Italy, and Russia	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD vs neurotypical children across one in-house and seven public fecal metagenomic datasets; in-house cohort included 64 ASD and 64 age-/sex-matched controls from the same Hong Kong catchment area.	n=452 (261 ASD, 191 neurotypical), median age 5 y in both groups. In-house n=128 (64/64). Diagnostic modeling reports discovery n=338 and random held-out validation n=79 (total 417); reason the remaining 35/452 were not in this modeling	Stool	Shotgun metagenomics; in-house Illumina NovaSeq 6000, PE150; all cohorts reprocessed through a standardized pipeline.	Microbial genomic structural variants (variable and deletion SVs) plus taxonomic species abundance.	12-marker SVM panel selected by sequential floating forward selection: 6 variable SVs + 3 deletion SVs + 3 bacterial species.	Discovery cohort (random 70% pooled across source cohorts): AUROC 0.858. Random held-out 30% from the same pooled multicohort source: AUROC 0.811. Single-type panels: v-SV 0.791, d-SV 0.752, species 0.723. Ten-fold CV used in algorithm/feature selection within discovery data.	Not a newly recruited external cohort: validation was a random holdout from the same eight source cohorts. Cross-by-cohort validation without retraining/re-optimization gave median AUROC 0.720 (Q1 0.673, Q3 0.730), providing transportability evidence

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							denominator is Unclear in the source report.						across source cohorts. Humanized-mouse validation was mechanistic, not clinical classifier validation.
71	Wan, 2024	39468837	Hong Kong SAR, China	Neuro/psych: Autism spectrum disorder	Disease detection	Clinician-confirmed ASD vs unrelated typically developing (TD) children. Discovery: cases from two Hong Kong teaching hospitals and community TD controls; validation: separate HATCH cohort with independently checked identities.	n=598 total based on detailed cohort counts: discovery n=273 (129 ASD, 144 TD), median age 6 [IQR 5-8] y; independent validation n=325 (135 ASD, 190 TD), median age 9 [7-10] y. Thus overall ASD=264, TD=334. Abstract wording "ASD vs TD=273 vs325" conflicts with these detailed counts.	Stool	Shotgun metagenomics; Nextera DNA Flex libraries; Illumina NextSeq 550, PE150.	Taxonomic markers plus metagenome-assembled genomes (MAGs; bacterial and viral).	Random forest with 49 markers: 5 bacterial taxa + 44 MAGs (42 bacterial, 2 viral). Direct feature-layer comparison (Figure 1): the combined taxonomy–MAG model achieved a mean AUROC of 0.858, compared with 0.809 for species-level taxonomy and 0.827 for MAGs alone using random forest classification.	Discovery cohort: repeated stratified 5-fold CV within 70% training workflow; AUROC=0.886 (95% CI 0.868-0.892), sensitivity 0.773, specificity 0.834.	Independent participant validation cohort (same Hong Kong region, identities checked independent): AUROC=0.734 (95% CI 0.682-0.782), sensitivity 80.0%, specificity 71.6%. No geographic external validation beyond Hong Kong.
72	Kang, 2017	29274915	United States (Arizona)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD vs neurotypical children; participants recruited in Arizona. Three neurotypical and six ASD participants had been in a prior cohort but provided fresh samples and new questionnaires.	n=44 (23 ASD, 21 neurotypical), age 4-17 y. Age mean ± SD: ASD 10.1 ± 4.1 y; neurotypical 8.4 ± 3.4 y.	Stool	Fecal metabolomics by 1H-NMR (Varian Direct Drive/VNMRS 600 MHz); bacterial 16S rRNA V2-V3 amplicon pyrosequencing (104F/530R), Roche Genome Sequencer FLX-Titanium.	Metabolites plus taxonomic/community microbiome features.	Five-metabolite Fisher discriminant analysis panel: caprate, nicotinate, glutamine, thymine, aspartate.	Leave-one-out CV: at >20% ASD-membership threshold, sensitivity 78% (18/23) and specificity 81% (17/21); at >30% threshold, sensitivity 70%, specificity 95%. Panel AUC not reported. Feature-combination search was performed in this small cohort; nesting of all selection steps within LOOCV is not fully clear.	No independent/external validation.
73	Peralta-Marzal, 2024	38191575	United States (discovery); China (two validation cohorts; secondary reanalysis of published datasets)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD vs neurotypical controls. Discovery controls were siblings of ASD cases; validation controls were unrelated age-matched neurotypical children. ASD diagnosis used ADOS in discovery and CARS or ADI-R/CGI-S/ABC-I in validation datasets.	Discovery n=117 (60 ASD, 57 siblings), mean age 5.02 ± 1.59 vs 4.56 ± 1.88 y. Validation 1 n=127 (77 ASD, 50 controls), 3.21 ± 0.98 vs 3.58 ±	Stool (secondary analysis)	16S rRNA amplicon sequencing reanalysis: discovery V4 paired-end 150 bp; validation cohorts V4 single-end 250 bp and V3-V4 paired-end 300 bp; DADA2/IDTAXA with SILVA r138.	Taxonomic (ASV-level) microbial profiles.	REFS selected 26 ASVs in the sibling-controlled discovery cohort; classification evaluated with multiple algorithms. In validation cohorts, corresponding transferred ASVs were evaluated with five classifiers.	Nested 10-fold CV in discovery: average AUC=0.816 for the 26-ASV signature; best classifier (MLP) AUC=0.90.	Two independent participant cohorts were used for feature transfer, but classifiers were re-fitted/evaluated by 10-fold CV within each validation cohort rather

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							1.21 y. Validation 2 n=96 (48/48), mean age about 5 vs 4 y (SD NR). Included ages 2-7 y.						than applying one locked discovery model. Average AUCs=0.748 and 0.740; best Extra Trees AUC=0.84 in each.
74	Chen, 2026	41421350	China (Shanghai/Suzhou in-house cohorts) plus eight public ASD cohorts from four countries/regions	Neuro/psych: Autism spectrum disorder	Disease detection	ASD children with recurrent GI symptoms, DSM-5 diagnosis confirmed with CARS/ABC, vs age- and gender-matched neurotypical children without GI symptoms. Discovery and independent in-house validation were hospital/community recruited; public metagenomic cohorts were used for transferability.	Discovery n=201 (102 ASD, 99 TD), age 3-10 y. Independent in-house validation n=69 (49 ASD, 20 TD). Eight public ASD cohorts contributed 853 fecal metagenomes for external testing; disease-specificity cohorts were also evaluated.	Stool	Shotgun metagenomics (Illumina NovaSeq 6000) plus untargeted fecal LC-MS metabolomics in in-house cohorts; public metagenomes reprocessed through a unified pipeline.	Multi-omics: taxonomic species, microbial genes/pathways, genomic variants (SVs/InDels/SNVs), and fecal metabolites.	Random-forest/xMarkerFinder models. Microbiome-only 20-marker panel: 13 species, 3 genes, 4 genomic variants. Integrated 20-marker multi-omics panel: 6 species, 1 microbial gene, 1 pathway, 1 SNV, and 11 metabolites.	Stratified 5-fold CV: five-layer microbial model AUC=0.962; 20-marker microbiome panel AUC=0.961. Metabolomics-only AUC=0.97; full integrated multi-omics AUC=0.98; selected 20-marker multi-omics panel AUC=0.99 in the training cohort.	Independent in-house validation: 20-marker multi-omics AUC=0.90; microbiome-only 20-marker AUC=0.890. The 20-marker microbiome model was additionally applied across the in-house validation cohort plus eight external public ASD cohorts: median AUC=0.78 (range 0.65-0.99). Public cohorts did not provide paired metabolomics, so they validate the microbiome panel, not the full multi-omics panel.
75	Basic-Cicak, 2024	39594202	Bosnia and Herzegovina (primary cohort); Italy (external published validation dataset)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD vs neurotypical children in Bosnia and Herzegovina; an independent published Italian pediatric ASD/healthy dataset was used for validation.	Primary n=60 (44 ASD, 16 controls), age 3-12 y; mean age conflict: Methods 6.2 vs 5.7 y, Results/Table 1 6.10 vs 5.25 y. External n=25 (11 ASD, 14 controls), age 2-4 y. SMOTE expanded the primary modeling data to 88 synthetic/balanced observations (44 per class), not 88 participants.	Stool	16S rRNA V3-V4 amplicon sequencing, Illumina MiSeq v3; predicted KEGG pathways from 16S profiles.	Taxonomic (genus-level) and predicted functional pathways.	RF, support-vector classifier, gradient boosting, and Extra Trees models using genus abundance or KEGG pathways; SMOTE and GridSearchCV used for primary modeling.	Five-fold CV on SMOTE-balanced modeling data: highest mean accuracy 80% for genus-level RF and Extra Trees; RF KEGG-pathway model accuracy 78%.	Independent Italian dataset (11 ASD/14 controls): the developed genus-level models were applied without re-training on the validation data; best mean accuracy 79% (Extra Trees). AUC was not reported for the external test.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
76	Su, 2024	38977906	Hong Kong SAR (discovery, hospital, and community cohorts); external public datasets from Asia, Europe, and America	Neuro/psych: Autism spectrum disorder	Disease detection	Children with ASD vs neurotypical children. Discovery cohort was clinically diagnosed; independent hospital and community cohorts were used for validation. ADHD and atopic-dermatitis cohorts were used for specificity testing.	Total n=1, 627 children, age 1-13 y. Discovery 709 ASD/374 NT, age 3-12 y. Independent hospital 82/90, age 4-11 y. Independent community 116/60; Results/Fig.1 report age 1-8 y whereas Methods report 3-12 y. External public data n=237 (121 ASD/116 NT), age 2-13 y.	Stool	Shotgun metagenomic sequencing with standardized processing; multikingdom taxonomic and functional profiling.	Multikingdom taxonomic (archaea, bacteria, fungi, viruses) plus microbial genes and metabolic pathways.	Random-forest ensemble; final 31-marker multikingdom/functional panel. Altered features included 14 archaea, 51 bacteria, 7 fungi, 18 viruses, 27 genes, and 12 pathways; model contribution was strongly driven by ubiquinol-7 and thiamine-diphosphate biosynthesis pathways.	Discovery: 31-marker panel AUC=0.91; single-kingdom panels AUC=0.68-0.87.	Independent hospital cohort: AUC=0.87, sensitivity 91%, specificity 73%. Independent community cohort: AUC=0.89; AUC=0.91 in age ≤6 y and ≤4 y subsets. Six public datasets (n=237): AUC=0.78, sensitivity 65.30%, specificity 72.4%.
77	Vernocchi, 2023	38192296	Italy (Rome)	Neuro/psych: Autism spectrum disorder	Disease detection; Phenotyping/stratification	ASD diagnosed by DSM-5 and confirmed by ADOS-2/ADI-R vs age-matched neurotypical controls. Within ASD, VOC profiles were explored by age, GI symptoms, ASD severity, and behavioral-risk scales.	n=76 (41 ASD, 35 controls). ASD age 3-15 y; 36 boys/5 girls. Controls were age-matched (21 boys/14 girls).	Stool	Fecal GC-MS/SPME volatile-organic-compound profiling plus 16S rRNA metataxonomic profiling.	Metabolites/VOCs with taxonomic/community context.	HistGradientBoostingClassifier using nine VOCs (including methyl isobutyl ketone, benzeneacetaldehyde, phenyl ethyl alcohol, ethanol, butanoic acid, octadecane, acetic acid, skatole, and tetradecanal); logistic-regression analyses supported overlapping VOC predictors.	Same-cohort HistGradientBoostingClassifier accuracy=89% for ASD vs controls. The source report does not clearly report a prespecified holdout or cross-validation scheme for this headline accuracy; therefore validation method is Unclear.	Not evaluated: no independent clinical cohort. Within-ASD age/GI/severity analyses are cross-sectional stratification/association, not external validation or prospective prediction.
78	He, 2023	37925571	China (southern China/Xiamen-based study)	Neuro/psych: Autism spectrum disorder	Disease detection	Children with constipated ASD (C-ASD) vs typically developing children without GI disorders. Thus, the classifier targets a GI-defined ASD phenotype and may capture constipation-related as well as ASD-related signals.	n=80 (40 C-ASD, 40 TD). Mean age: C-ASD 5.30±1.34 y vs TD 5.83±1.28 y (P=0.077).	Stool	16S rRNA V3-V4 amplicon sequencing (Illumina HiSeq 2500, PE250) plus fecal SCFAs by GC-MS (Agilent GC with 5977B MS).	Taxonomic genera plus metabolites/SCFAs.	Random forest based on 16 discriminant genera; propionate-only model; and combined 16-genera + propionate model.	Five-fold CV: 16-genus model AUC=0.924 (95% CI 0.896-0.953); propionate-only AUC=0.620 (0.565-0.675); combined model AUC=0.924 (0.895-0.952), showing no improvement over taxa alone.	No independent/external validation.
79	Zhao, 2023	37271221	China (Beijing)	Neuro/psych: Autism spectrum disorder	Disease detection	Children aged 3-6 y with ASD plus GI symptoms vs age-/sex-matched typically developing children without GI symptoms. ASD was DSM-5/ADOS diagnosed and had CARS >30 plus at least one clinically relevant 6-GSI symptom.	n=76 (36 ASD+GI, 40 TD). Mean age 3.92 ± 1.08 vs 4.30 ± 1.16 y. Model performance was evaluated in another 12-person population (5 ASD, 7	Stool	16S rRNA V3-V4 amplicon sequencing; DADA2/SILVA-based ASV profiling. Exact sequencing instrument is NR in the source report.	Taxonomic genera plus predicted functional KEGG/gut-metabolic/gut-brain modules.	Genus-level support-vector machine with CLR-transformed abundances; SHAP used for feature interpretation.	Development performance was not separately quantified in the source report.	Validation population n=12 (5 ASD, 7 TD): AUC=0.88. The paper calls this another/validation population, but whether it was a separately recruited

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							TD).						independent cohort or a prespecified holdout from the 76 participants is not clearly reported; independence is Unclear.
80	Wang, 2023	36838256	China (Shenzhen) and Russia (Moscow); secondary cross-cohort analysis	Neuro/psych: Autism spectrum disorder	Disease detection	Public pediatric ASD vs typically developing stool metagenomes from Shenzhen and Moscow; no new participants were recruited.	Total n=124 (73 ASD, 51 TD). Shenzhen n=74 (43 ASD, 31 TD), age 2-7 y; Moscow n=50 (30 ASD, 20 TD), age 3-5 y.	Stool (secondary analysis)	Public shotgun metagenomic datasets reprocessed with a common species-profiling pipeline (MetaPhlAn2); original sequencing platforms differed by cohort.	Taxonomic species-level profiles.	Random-forest iterative feature-selection/prediction pipeline using microbial species; models were evaluated separately in each cohort and in pooled data.	Repeated stratified 10-fold CV: Shenzhen AUROC=0.984, accuracy=97%; Moscow AUROC=0.81, accuracy=67%; pooled cohorts AUROC=0.86, accuracy=80%.	Not evaluated as a locked external model: both source cohorts were used within cross-validated model development/analysis rather than one cohort training a fixed classifier that was directly tested in the other.
81	Lou, 2022	34930815	China (25 provinces; mainly Hunan, Shandong, Zhejiang, and Guangdong); two additional published Chinese validation cohorts	Neuro/psych: Autism spectrum disorder	Disease detection; Disease association	Clinically defined ASD vs neurotypical participants across age strata; analyses also related microbiome development and microbial features to ASD severity, GI symptoms, sleep, allergy, and age.	Main text: 1, 222 participants, 773 ASD (16 mo-19 y), 429 NT children (11 mo-15 y), plus 20 healthy adults (16-24 y); this statement is internally inconsistent arithmetically and with Supplementary Table 2 (current cohort 772 ASD/450 NT including adults). Validation 1 n=105 (73 ASD/32 NT); validation 2 n=286 (143/143); validation 3 n=86 (45/41).	Stool	16S rRNA V4 amplicon sequencing, Illumina NovaSeq 6000, paired-end 2x150 bp; predicted gut-brain modules and MetaCyc functions.	Taxonomic OTUs/community plus predicted microbial metabolic functions.	Age-aware random-forest/SHAP classifier using a top-20 feature set; the abstract highlights 2 OTUs (Veillonella and Enterobacteriaceae) plus 17 microbial metabolic functions, with physiological age also incorporated into performance evaluation.	Current-cohort AUC=0.86. Age-stratified and feature-level analyses were exploratory; exact model-locking details are described across main/supplementary methods.	Validation cohort 1 (73 ASD/32 NT) AUC=0.78; independent published validation cohort 2 (143/143) AUC=0.82; independent published validation cohort 3 (45/41) AUC=0.67. Performance decline across cohorts indicates limited transportability.
82	Wan, 2022	34312160	Hong Kong SAR	Neuro/psych: Autism spectrum disorder	Disease detection	Community-recruited preschool children with clinically diagnosed ASD vs age- and gender-matched typically developing children. Independent validation children came from different community sources in Hong Kong.	Discovery n=128 (64 ASD, 64 TD), age 3-6 y; median age 59 [51-66] vs 56 [44-66] months. Independent	Stool	Deep shotgun metagenomic sequencing, Illumina NovaSeq 6000; species and microbial functional profiling.	Taxonomic species plus functional/age-maturity features.	Random-forest five-species panel: Alistipes indistinctus, candidate division TM7 isolate TM7c, Streptococcus cristatus, Eubacterium limosum, and	Discovery: five species selected by 10-fold CV; AUC=0.826 for ASD vs TD.	Independent community validation using the same five biomarkers: AUC=0.762. The article confirms an independent

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							validation n=18 (8 ASD, 10 TD). Total n=146.				Streptococcus oligofermentans.		cohort; whether the final RF coefficients/threshold were completely locked versus re-fitted on the five-marker set is not explicit.
83	Ding, 2020	32912596	China (Zhengzhou)	Neuro/psych: Autism spectrum disorder	Disease detection; Phenotyping/stratification	Children with DSM-5 ASD vs unrelated age-/sex-matched healthy controls; ASD severity was further stratified by CARS into mild ASD (30<CARS<36) and severe ASD (CARS ≥36).	n=127 (77 ASD: 33 mild, 44 severe; 50 HC), age 2-7 y. Mean age 38.5 ± 11.7 vs 42.9 ± 14.5 months.	Stool	16S rRNA V4 amplicon sequencing, Illumina HiSeq paired-end 250 bp.	Taxonomic species/genus-level microbial profiles and community diversity.	Random-forest classifier. Detailed Results identify a 10-species predictive signature; the Abstract instead highlights three genera (Faecalitalea, Caproiciproducens, Collinsella), creating a model-description inconsistency.	Random 80%/20% training/testing split. Detailed Results: 10-species model training AUC=0.994 and internal testing AUC=0.983. Abstract reports a three-marker RF AUC=0.94 and validation AUC=0.98 (95% CI 97.9%-100%), so exact correspondence of the abstract model to the detailed 10-species model is Unclear.	Not evaluated: no independent/externally recruited cohort; the 20% test set is an internal random split.
84	Zhai, 2019	30743242	China (Beijing, Jiangsu, and Shandong)	Neuro/psych: Autism spectrum disorder	Disease detection	Children meeting DSM-IV-TR/ICD-10 ASD criteria vs age- and region-matched healthy children without ASD/other neurologic disorders; controls also excluded chronic GI problems and family relationship to ASD.	n=136 for hair trace elements (78 ASD, 58 controls), mean age 4.96 ± 1.01 vs 4.90 ± 0.97 y. Only 88 fecal samples were available for microbiome analysis; case/control stool counts and classifier denominator are NR.	Stool; scalp hair	Stool 16S rRNA V3-V4 amplicon sequencing, Illumina MiSeq. Hair: ICP-MS (Pb/Cd/As/Cu/Zn/Fe), atomic fluorescence (Hg), and flame atomic absorption (Ca/Mg).	Host-microbe composite: taxonomic genera plus trace-element concentrations.	Random-forest models using the optimal 10 genera alone or combining microbial and element features.	Cross-validation was reported but the exact resampling scheme/test denominator was NR. Microbiota-only model: accuracy 64%, AUC=0.675. Combined microbial+metal model: accuracy 84%, AUC=0.903 on the reported testing data.	No independent/external validation.
85	Wang, 2021	33510860	China; public external ASD dataset for secondary validation	Neuro/psych: Autism spectrum disorder	Disease detection	Children aged 2-8 y with DSM-5 ASD vs age-matched typically developing children; participants with other major psychiatric diagnoses or recent antibiotics were excluded. External public metagenomes (SRP182132) were used to test VFGM markers.	Primary cohort n=74 (43 ASD, 31 TD); ASD age 2–6 y and TD age 2–5 y. External SRP182132 cohort n=77 (57 ASD, 20 TD); age distribution NR in this article.	Stool	Shotgun metagenomic sequencing (Illumina HiSeq 4000) plus fecal IgA by ELISA.	Functional/virulence-factor microbial genes plus host-microbe immune composite (fecal IgA and VFGM diversity).	Random-forest models using virulence-factor genes. Primary candidates included GBS cpsH/cpsJ/cpsO, VFGM Shannon diversity, and fecal IgA.	Primary cohort: random 60%/40% train/test split with 5-fold CV for AUC uncertainty. Three GBS genes AUC=0.974; three GBS genes + IgA + VFGM diversity AUC=0.958; top-10 VF model AUC=1.0. Figure/text mapping of 0.974 vs 0.958 is internally inconsistent.	Public SRP182132 analysis required re-selecting a different set of GBS/VF genes rather than directly applying the primary locked panel: top-3 new GBS genes AUC=0.833; top-3 GBS +2 non-GBS VF genes AUC=0.92; 5 GBS +4 non-GBS genes

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
													AUC=0.929. This is external candidate-feature/model redevelopment, not direct validation of the primary composite classifier.
86	Liu, 2021	34233671	China (Shijiazhuang, Hebei)	Neuro/psych: Pediatric myasthenia gravis	Disease detection	Treatment-naive pediatric MG (all MGFA class I/ocular MG) confirmed by clinical presentation plus at least one supportive diagnostic test vs age-/gender-matched healthy controls. New MG cases were recruited for validation.	Discovery n=99 (53 MG, 46 HC), mean age 6.64 ± 3.79 vs 7.51 ± 3.79 y. Validation added 19 new MG cases; test set combined these 19 new cases with 12 HCs re-used from the discovery cohort.	Stool; serum (SCFA mechanistic assay)	Stool whole-genome shotgun metagenomics, Illumina HiSeq 2500; serum SCFAs by GC-MS (Thermo Trace1300-TSQ9000).	Taxonomic species panel; functional pathways/SCFAs were mechanistic supportive evidence.	Random-forest five-species panel: Fusobacterium mortiferum, Prevotella stercorea, Prevotella copri, Megamonas funiformis, and Megamonas hypermegale.	Cross-validation AUC=0.94 in the discovery cohort. The paper text describes fivefold CV, whereas the Fig.3 caption refers to five trials of 10-fold CV error; fold description is internally inconsistent.	Partially independent validation: test set contained 19 newly recruited MG cases plus 12 HCs randomly selected from the discovery cohort; AUC=0.84. Thus, cases were independent but controls were re-used, so this is not a fully independent external cohort.
87	Pang, 2025	39953112	China (Shanghai)	Immune/allergic: IgA vasculitis (Henoch-Schonlein purpura)	Disease detection	Children with clinically diagnosed HSP vs healthy children, with exclusion of infectious disease, major immune/metabolic conditions, oral disease, and recent probiotics/antibiotics/immunomodulators.	n=26 (13 HSP, 13 healthy controls). Mean age 6.39 ± 2.36 vs 7.14 ± 3.20 y.	Tongue-coating swab	16S rRNA V3-V4 amplicon sequencing (338F/806R), Illumina MiSeq 2x300 bp.	Taxonomic genera/community plus predicted microbial functions.	Random-forest classifier; five genera selected as the optimal biomarker set by 10-fold cross-validation.	Internal 10-fold CV/model ROC: AUC=0.790.	No independent/external validation.
88	Chen, 2025	40201970	China (Fujian; multicenter human recruitment)	Immune/allergic: IgA vasculitis (Henoch-Schonlein purpura)	Disease detection; Disease association	Relapsing IgA vasculitis children vs healthy controls. Human data were derived from a previous study; selected pooled human stools were additionally transplanted into mice to investigate mechanism.	Human n=39 (16 relapsing IgAV, 23 HC); median age 6.0 [5.0-9.0] vs 9.0 [6.2-11.4] y. For FMT, stools from 5 IgAV and 5 HC children were pooled separately for recipient mice.	Stool (human); murine stool/serum for mechanistic FMT experiments	Human/murine 16S rRNA V3-V4 amplicon sequencing, Illumina NovaSeq 6000 SP 2x250 bp; mouse serum cytokines/IgA by ELISA.	Taxonomic family-level microbiota; immune-cytokine mechanistic associations.	Random forest was used for feature ranking; human diagnostic ROC centered on Burkholderiaceae (predominantly Alcaligenaceae) abundance as a single microbial marker.	Human cohort ROC using Burkholderiaceae abundance: AUC=0.766, sensitivity=62.5%, specificity=91.3%. No train/test split or resampling-based validation was reported for this single-marker human ROC.	No independent clinical validation. The mouse FMT experiment demonstrates mechanistic transfer/association and must not be counted as external validation of a pediatric diagnostic classifier.
89	Tejesvi, 2021	33585283	Finland (Oulu)	Immune/allergic: PFAPA syndrome	Disease detection	Children undergoing tonsillectomy for PFAPA vs controls undergoing tonsillectomy for hypertrophied tonsils/non-infectious indications. PFAPA required recurrent patterned fever with asymptomatic intervals.	Analytic cohort n=52 (30 PFAPA, 22 controls). Median age at surgery 3.4 y (range 1.7-18.2) vs 5.6 y	Tonsil tissue	Fungal ITS2 amplicon sequencing (ITS7b/ITS4), Ion Torrent PGM with Ion 316 Chip; QIIME2/DADA2.	Taxonomic mycobiome/community.	Random-forest genus-level classifier.	Stratified 10-fold CV/out-of-bag analysis: AUC=0.47, SD=0.05, p=0.809 vs shuffled/random labels; no diagnostic discrimination above chance.	Not evaluated: no independent external cohort. The negative result remains a valid disease-detection study.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							(2.7-15.2); the case range includes at least one participant slightly >18 y and cannot be cleanly separated from the reported aggregate data.						
90	Qian, 2020	32264859	China (Zhejiang; Han Chinese)	Immune/allergic: Juvenile idiopathic arthritis	Disease detection; Disease association	Children with JIA vs age-, sex-, BMI-, and ethnicity-matched healthy controls. Clinical subtype/activity/laboratory indices were also correlated with microbial taxa, but no separate phenotype classifier was developed.	Enrolled n=82 (40 JIA, 42 HC), age 1-16 y; successfully sequenced n=81 (39 JIA, 42 HC). Median age 10.27 vs 9.95 y.	Stool	16S rRNA V3-V4 amplicon sequencing (341F/805R), Illumina MiSeq v3 600-cycle.	Taxonomic genus/community features.	Random forest with nested cross-validation selected 12 genera as biomarkers; a 12-genus RF disease classifier was evaluated by 10-fold cross-prediction and decision-curve analysis.	Internal cross-validation/cross-prediction: AUC=0.7975; genus-level RF predictive accuracy 67.9% in model-selection analysis. DCA showed net benefit over a threshold range reported in the paper.	No independent/external validation.
91	Xu, 2025	39369895	China (Chaozhou/Chaoshan, Guangdong)	Immune/allergic: Cow's milk protein allergy	Disease detection	Children with CMPA diagnosed according to the study's 2019 milk-allergy guidance vs healthy infants/young children selected during physical examination; all participants lived in the Chaoshan area.	n=46 (30 CMPA, 16 healthy controls). CMPA age strata: 0–14 mo (n=10), 15–30 mo (n=11), and >31 mo (n=9). Median age (IQR): 18.00 (21.00) vs 14.50 (24.00) mo.	Stool (feces)	16S rRNA V3-V4 amplicon sequencing; Illumina NovaSeq; QIIME2/DADA2.	Taxonomic (family/genus) and community/diversity.	Six-taxon panel: Peptostreptococcaceae; Clostridium within Peptostreptococcaceae; Clostridium within Lachnospiraceae; Coprococcus; Oscillospira; Eggerthella; random-forest analysis also reported.	Same-cohort exploratory ROC analysis: individual AUCs 0.6854-0.7271; six-taxon panel AUC=0.8708, P<0.0001. The source report does not clearly report an independent train/test split or cross-validation for this diagnostic AUC.	Not evaluated: no independent/externally recruited diagnostic cohort.
92	Sun, 2019	31431508	China (Beijing, Qingdao); United States (Denver, Colorado)	Immune/allergic: Atopic dermatitis/eczema	Disease detection; Phenotyping/stratification; Prognosis/treatment-response prediction	Mild-to-moderate active pediatric AD vs healthy site-/age-matched children across Beijing, Qingdao, and a previously published Denver cohort. Beijing AD children also underwent 4-week topical treatment; pretreatment MiSH was explored for response stratification.	Beijing: 28 AD + 30 healthy; Qingdao: 29 AD + 30 healthy; Denver: 59 AD + 13 healthy. Overall pediatric ages approximately 3-12 y (Beijing/Denver explicitly 4-12 y). Treatment-response stratification used 18 Beijing patients (type I n=7; type II n=11).	Skin swabs: lesional and nonlesional sites, mainly forearm/shank plus other body sites	Bacterial 16S rRNA V1-V3 (27F/534R). Beijing: Roche 454 FLX and Illumina MiSeq; Qingdao/Denver: Illumina MiSeq; computational processing harmonized.	Taxonomic (genus-level) and community/diversity.	Random-forest Microbial Index of Skin Health (MiSH) based on 25 genera; baseline MiSH also defined two treatment-response host types.	10-fold CV city models: Beijing AUC=0.97, accuracy=91.3%; Qingdao AUC=0.95, accuracy=89.4%; Denver AUC=0.78, accuracy=79.2%. Pooled three-city 25-genus MiSH: AUC=0.90, accuracy=86.4%. In the 18-patient treatment subset, type I vs II baseline MiSH 0.9 ± 1.1 vs 32.5 ± 11.5; delta-MiSH 29.2 ± 10.1 vs 15.9 ± 16.5 (P=0.032).	Geographic model transfer: Beijing-trained model -> Qingdao AUC=0.75, Denver=0.70; Qingdao-trained -> Beijing=0.89, Denver=0.73. No independent validation of the small treatment-response stratification.
93	Peng, 2026	42327796	China (Guangzhou, Guangdong)	Immune/allergic: Atopic	Disease detection; Disease association	Mild-to-moderate pediatric AD recruited at Guangdong Provincial Hospital of Chinese	n=69 (53 AD, 16 healthy), age 2-12 y.	Stool (feces)	16S rRNA V3-V4 (341F/806R), Illumina NovaSeq 6000; untargeted	Taxonomic/community plus fecal metabolites; diagnostic	Six microbial features: Bacteroidota, Parabacteroides, and	5-fold CV: combined panel AUC=0.941, accuracy=84.6%,	Not evaluated: no independent/external

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
				dermatitis/eczema		Medicine vs healthy community controls; recent systemic antibiotics, probiotics/prebiotics, or corticosteroids and major inflammatory/metabolic comorbidity were excluded.	Mean age: AD 6.47 ± 2.76 y; controls 6.12 ± 2.28 y.		fecal metabolomics by Vanquish UHPLC-HILIC coupled to Orbitrap Exploris 48 MS (ESI).	panel itself is taxonomic.	four tentative 16S species-level assignments (Bacteroides plebeius, B. thetaiotaomicron, B. xylanisolvens, Parabacteroides merdae); random forest.	sensitivity=90.0%. Feature selection/hyperparameter tuning was not nested; authors explicitly describe the model as exploratory. Severity correlations were association analyses, not a severity classifier.	cohort; article explicitly calls for independent external validation.
94	Fieten, 2018	29393195	Netherlands (Utrecht)	Immune/allergic: Atopic dermatitis with/without food allergy	Phenotyping/stratification	Children with AD (UK Working Party criteria) with confirmed food allergy vs AD without food allergy. Food allergy was defined by positive DBPCFC or convincing immediate clinical history plus sensitization; peanut could also be classified using the study's Ara h 2 threshold.	n=82 children with AD: 20 food allergy, 62 no food allergy; overall median age 2.5 y. Food-allergy group median 2.2 y (range 0.5-12); no-food-allergy group 3.0 y (0.5-19).	Stool (feces)	16S rRNA V4 amplicon sequencing (adapted 533F/806R); Illumina MiSeq; mothur.	Taxonomic (species-level).	Elastic-net six-species signature: Bifidobacterium breve, B. pseudocatenulatum, B. adolescentis, Escherichia coli, Faecalibacterium prausnitzii, and Akkermansia muciniphila.	Internal random holdout/cross-validation by hiding 30% of children: AUC=0.83, sensitivity=0.77, specificity=0.80.	No independent/external validation.
95	Jiang, 2022	34997172	South Korea (Seoul)	Immune/allergic: Atopic dermatitis/eczema	Disease detection	Children with AD diagnosed by Hanifin-Rajka criteria at Asan Medical Center vs healthy controls; moderate-to-severe AD was represented in the source cohort.	n=161 (88 AD, 73 controls), age 6-72 mo. Mean age: AD 17.37 ± 3.48 mo; controls 10.81 ± 2.15 mo.	Feces: gut microbiota plus exfoliated colonocytes recovered from fecal specimens	Host transcriptome: Affymetrix GeneChip Human Gene 2.0 ST Array; gut microbiota: 16S rRNA V1-V3 using Roche/454 FLX Titanium and Illumina MiSeq.	Host-microbe composite: colonocyte transcriptome plus taxonomic gut microbiota.	SVM with radial-basis kernel; selected 35 transcriptome features plus 50 microbiota features (4 phyla, 46 genera; 85-feature combined classifier).	Stratified split: training n=131, held-out test n=30. Combined model: F1=0.7778, accuracy=0.7333, precision=0.7000, recall=0.8750; AUC=0.75. Test-set isolation is imperfect because min-max normalization was performed on training and test sets together.	Not evaluated: no independent/external cohort; replication in other cohorts was stated as needed.
96	Zhang, 2019	30860574	China (Beijing)	Immune/allergic: Atopic dermatitis/eczema	Disease detection	Infants/young children with pediatrician-diagnosed eczema vs healthy children from the same clinic; eczema required at least two prespecified clinical signs.	Total recruited n=229, age <3 y: healthy n=123, discovery eczema n=49, new eczema qPCR cohort n=57. Mean ages: 1.28 ± 0.90, 1.74 ± 1.34, and 1.25 ± 0.88 y, respectively. Sequencing discovery set n=172 (123 healthy, 49 eczema).	Stool (feces)	16S rRNA V3-V4 amplicon sequencing; Illumina HiSeq 2500; targeted qPCR for Bifidobacterium/Bacteroides on ABI 7500.	Taxonomic (genus/OTU) and community/diversity.	Random-forest ranking of genera selected after differential-abundance testing; Bifidobacterium was the principal candidate marker. Targeted qPCR was used for technical/candidate-marker confirmation.	Same discovery cohort: Bifidobacterium OTUs AUC=0.83; Enterobacter=0.79, Roseburia=0.49, Erwinia=0.24. A separate resampling scheme for these ROC values was not reported.	Candidate-biomarker confirmation only: a new eczema cohort (n=57) showed lower Bifidobacterium by qPCR versus the healthy reference group (P=0.0031). No independent fixed classifier performance was reported.
97	Zhang, 2025	4010856	China (Zhuhai, Guangdong)	Immune/allergic: Food	Disease detection;	Oral-food-challenge-defined food allergy with	n=120, age 1-6 y:	Fresh stool (feces)	16S rDNA V3-V4 reported by	Taxonomic (genus-level) and	Random forest identified 14 pivotal	Same-cohort internal split: 60%	Not evaluated: no

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
		1		allergy with/without malnutrition	Phenotyping/stratification	adequate weight (FANM) or malnutrition (FAM) vs age-/sex-matched healthy children without allergy or malnutrition; FAM also stratified by malnutrition severity.	FAM n=40, FANM n=40, healthy n=40. Mean age: 3.00 ± 1.64, 2.65 ± 1.44, and 2.75 ± 1.49 y, respectively; FAM included moderate n=19 and severe n=21 malnutrition.		authors (515F/806R); paired-end Illumina NovaSeq 6000.	community/diversity.	genera; probability-of-disease (POD) classifiers built separately for FAM vs healthy and FANM vs healthy, with covariate-adjusted analyses.	training / 40% testing after RF biomarker selection. Reported AUC: FAM vs healthy 0.8986 (95% CI 0.7640-0.9973); FANM vs healthy 0.8892 (95% CI 0.7358-0.9965). No direct FAM-vs-FANM classifier was reported.	independently recruited/external cohort.
98	Li, 2021	34222048	China (Qingdao, Shandong; Guangzhou, Guangdong)	Oral: Dental caries	Disease detection	Urban primary-school children aged 6-8 y with mixed dentition: caries-affected vs caries-free; Qingdao and Guangzhou cohorts were recruited independently by city.	n=130: Qingdao n=96 (48 caries, 48 healthy); Guangzhou n=34 (17 caries, 17 healthy); age 6-8 y.	Unstimulated saliva	16S rRNA V1-V3 amplicon sequencing; Roche 454 GS FLX Titanium.	Taxonomic (species-level) and community/diversity.	Random-forest 14-species caries panel after accounting for geography.	10-fold CV: Qingdao AUC=0.9102 (95% CI 0.8527-0.9676); Guangzhou AUC=0.8616 (0.7157-1.0000); pooled two-city AUC=0.9217 (0.8745-0.9688), accuracy=83.08%.	Cross-city model transfer: Qingdao-trained -> Guangzhou AUC=0.87; Guangzhou-trained -> Qingdao AUC=0.89. These are geographic holdouts, but the final 14-species feature set was selected using the combined two-city analysis; therefore not a fully independent locked external validation.
99	Zhang, 2021	34035879	China (Qingdao, Shandong)	Oral: Early childhood caries	Disease detection	Children aged 4-6 y with active ECC vs caries-free controls from the same population.	n=331: 165 caries-active and 166 caries-free; age 4-6 y.	Unstimulated saliva	Microbiome: 16S rRNA V3-V4 amplicons, Roche 454 FLX Titanium; saliva chemistry: ion chromatography for electrolytes plus pH/total protein; qPCR technical confirmation of selected taxa.	Taxonomic salivary microbiota plus biochemical/electrolyte host-fluid features.	Random-forest classifiers built separately from microbial features, electrolyte/pH features, and combined microbial-biochemical features.	Internal same-cohort resampling/holdout: microbiome-only AUROC=0.70 (95% CI 0.6386-0.7515), accuracy=0.65; electrolyte-based AUROC=0.94 (0.9150-0.9691), accuracy=0.89; combined AUROC=0.89 (0.8581-0.9280), accuracy=0.82.	No independent/external validation.
100	Wang, 2019	31690590	China (Lin'an, Zhejiang)	Oral: Severe early childhood caries	Disease detection; Prognosis/treatment-response prediction	Preschool children with severe ECC (dmfs ≥8) vs caries-free children (dmfs=0). Baseline healthy children were followed for 6 months and classified as remaining healthy or developing caries.	n=44, age 3-5 y: severe ECC n=25; healthy n=19. Among baseline healthy children, 15 remained healthy and 4 developed new caries	Saliva	Shotgun metagenomic sequencing; Illumina HiSeq 2000; species/strain profiling and direct functional gene/pathway analysis.	Taxonomic (species/strain) plus functional genes/pathways.	mRMR-selected seven-species SVM: Streptococcus mutans, Prevotella amnii, Eubacteriaceae bacterium ACC19a, Shuttleworthia satelles, Dialister invisus, 'Candidatus Nitrospira defluvii', and an Erysipelotrichaceae	Current severe-ECC vs healthy classifier: AUC=0.9853 (95% CI 0.9581-1.0000); feature-subset selection used leave-one-out CV. Future-onset analysis among 19 initially healthy children: AUC=0.7833 for H2C (n=4) vs	Not evaluated: no independent/external cohort; the prospective-onset analysis was within the same small cohort (4 incident cases).

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
							by 6 months.				bacterium.	H2H (n=15).	
101	Hart, 2011	22013442	Brazil (Montes Claros, Minas Gerais)	Oral: Early childhood caries	Disease detection	High-risk, low-socioeconomic-status children with little routine dental care: caries-active vs caries-free by clinical dental examination.	n=204, age 1-8 y (overall 3.83 ± 2.55 y): caries-free n=118, mean age 2.3 ± 0.2 y; caries-active n=86, mean age 6.02 ± 0.2 y. Microbial analysis used 448 plaque samples (118 caries-free; 330 lesion/site samples from caries-active children).	Supragingival dental plaque; saliva	Plaque microbiota: reverse-capture checkerboard hybridization targeting 82 oral bacterial species/groups; salivary proteomics: SELDI-TOF-MS.	Targeted taxonomic plus salivary proteomic host-microbe composite.	Linear SVM, Wilcoxon feature-restricted SVM, and random forest; combined microbial plus CM-10 low-intensity salivary proteomic features.	Repeated internal random subsampling with 70/30 train/test split across 40 splits. Best combined feature-restricted SVM: test error 6.00% ± 2.67%, sensitivity 92.68% ± 4.46%, specificity 95.20% ± 3.87%; ROC AUC=0.96 ± 0.02.	No independent/external validation. The report does not clearly state whether repeated plaque samples from the same child were grouped within the same train/test partition.
102	Wei, 2025	40676575	Public multi-study secondary analysis: China, United States, United Kingdom, Denmark, Australia, and Taiwan	Oral: Dental caries	Disease detection	Public oral-microbiome samples from 10 prior studies were reprocessed and grouped by age, matrix, and sequencing method; caries vs healthy status came from source studies. The pooled evidence includes pediatric and adult cohorts.	n=576 samples from 10 studies: 496 shotgun metagenomes (251 plaque, 245 saliva) plus 80 full-length 16S samples. Pediatric, adolescent, and adult age groups are mixed; under-6 subgroups were modeled separately.	Saliva; dental plaque (secondary analysis)	Reprocessed public shotgun metagenomics (MetaPhlAn/HUMAnN) and full-length 16S rRNA datasets (QIIME2/DADA2), according to source data.	Taxonomic species plus functional pathways.	Seven architectures (logistic regression, RF, SVM, gradient boosting, CNN, feedforward neural network, Transformer) with subgroup-specific routing/ensemble learning.	Random 80% training / 20% held-out test, repeated 10 times within each data group. Under-6 saliva: RF/GB/FNN reached AUC=1.00 and 100% accuracy; under-6 plaque: FNN/Transformer AUC=1.00 and 100% accuracy. Full-length 16S maximum AUC=0.93, accuracy=87.5%.	Not evaluated as independent external validation: source studies were pooled and randomly split at sample level rather than held out by study/cohort; authors state entirely independent-cohort validation remains needed.
103	Lee, 2021	34441259	South Korea (Yangsan)	Oral: Dental caries	Disease detection	Children visiting Pusan National University Dental Hospital; non-caries = ICDAS 0-2 only, caries = any ICDAS ≥3. Models were stratified by age group and oral matrix.	n=120. Group 1 (<6 y): n=60, mean 4.5 ± 0.7 y (range 3.1-5.9), 30 caries/30 non-caries. Group 2 (6-12 y): n=60, mean 7.9 ± 1.3 y (6.0-11.8), 30 caries/30 non-caries.	Saliva; dental plaque (paired from each child)	16S rRNA V3-V4 amplicon sequencing; Illumina HiSeq.	Taxonomic (genus/species-level).	Random forest using differential taxa; separate models for <6-y plaque/saliva and 6-12-y plaque/saliva.	Stratified 5-fold CV. Accuracy: <6 plaque 0.78 ± 0.14; <6 saliva 0.70 ± 0.18; 6-12 plaque 0.73 ± 0.10; 6-12 saliva 0.83 ± 0.13 (balanced accuracy 0.84, precision 0.87, sensitivity 0.82, specificity 0.86). AUC was not reported.	No independent/external validation.
104	Teng, 2015	26355216	China (Guangzhou, Guangdong)	Oral: Early childhood caries	Disease detection; Prognosis/treatment-response prediction	Longitudinal preschool cohort followed for 2 y: children remained healthy, transitioned from health to caries, or had caries	Longitudinal cohort n=50, baseline approximately 4 y (51	Supragingival plaque; saliva	16S rRNA V1-V3 amplicon sequencing; Roche 454 FLX Titanium.	Taxonomic/community (species-level MiC); age-detrended microbial	Random-forest Microbial Indicator of Caries (MiC), with plaque, saliva, and composite models;	Subject-aware internal CV was used to avoid placing the same child in training and test partitions;	Independent 40-child saliva cohort for disease detection: AUC=0.72. No

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						progression. An additional independent saliva cohort comprised severe ECC and healthy children.	± 6 mo); 284 samples (142 plaque, 142 saliva) across 4 time points. Groups: H2H n=17, H2C n=21, C2C n=12. Independent detection cohort n=40, age 46-50 mo (20 healthy, 20 severe ECC).			signatures.	simplified eight-Prevotella model also examined.	composite species MiC AUC=0.77 (summary diagnostic accuracy 70%). For future onset, plaque MiC classified 17/21 (81%) pre-disease RelativeH samples as caries; saliva 11/21 and composite 15/21.	independent longitudinal cohort validated the future-onset prediction.
105	Sun, 2023	37056188	Global public microbiome datasets; optimized pediatric subsets from China	Oral: Dental caries	Disease detection	Secondary analysis of public human oral 16S datasets from NCBI, MG-RAST, EMBL-EBI, and QIITA. A large healthy reference database was compared with caries/healthy test data; host-factor-controlled analyses focused on Chinese pediatric plaque subsets.	Reference database: 20, 540 healthy oral microbiome samples. Test dataset n=1, 703 (860 healthy controls, 843 caries). Optimized Chinese pediatric plaque datasets: high-caries n=149, medium n=159, low n=206; mixed-plaque dataset n=500.	Heterogeneous oral samples in public data (mainly saliva and dental plaque; also other oral sites); pediatric optimized analyses used dental plaque	Public 16S rRNA amplicon datasets spanning heterogeneous regions/platforms; microbial search engine (MSE)/Meta-Storms with Greengenes-based taxonomic normalization.	Community/taxonomic similarity expressed as Microbiome Novelty Score (MNS).	MSE-derived MNS caries classifier; host-factor optimization for country, age, DMFT, and sampling site.	Overall retrospective test dataset AUC=0.67. After host-factor restriction/optimization, Chinese pediatric plaque AUCs: high caries 0.87; medium 0.74; low 0.74; mixed plaque 0.75.	No newly recruited independent cohort. Public multi-project testing demonstrates retrospective cross-dataset transportability, but reference/test construction occurred within one secondary database framework rather than a prospectively locked external validation.
106	Zhang, 2026	42320590	Hong Kong SAR, China	Oral: Early childhood caries	Phenotyping/stratification	All participants had ECC with at least one active cavitated lesion. Each child contributed paired supragingival plaque from a carious lesion and a spatially matched intact tooth; lesions were further staged as enamel vs dentine caries.	n=84 children, age 3-4 y; 168 paired plaque samples (84 carious, 84 matched intact). Caries stage: enamel ICDAS 3 n=58; dentine ICDAS ≥4 n=26; ECC n=57 and severe ECC n=27.	Site-specific supragingival dental plaque from carious lesion and matched intact tooth	16S rRNA amplicon sequencing (341F/806R); Illumina MiSeq PE300; PICRUSt2 for inferred functional prediction.	Taxonomic species plus inferred functional pathways.	Logistic-regression ROC classifiers from candidate species: 13-species caries-vs-intact panel; stage-specific six-species panel.	5-fold CV: 13-species panel AUC=0.85 (95% CI 0.79-0.91) for carious vs matched intact sites. Stage-specific six-species panel: AUC=0.89 (0.81-0.98) for dentine-stage caries vs matched controls; AUC=0.81 (0.69-0.92) for dentine- vs enamel-caries lesions.	No independent/external validation. The CV split unit was not reported, and candidate-feature discovery appears to precede CV; paired sites from the same child therefore require cautious interpretation.
107	Yang, 2025	40449489	China (Qingdao, Shandong)	Oral: Early childhood caries	Disease detection; Prognosis/treatment-response prediction	Two longitudinal preschool cohorts tracked individual teeth. Cohort B was used for spatiotemporal model development; cohort A was an independent participant cohort used for in-study validation. Teeth could remain	n=89 preschoolers, baseline 52 ± 10 mo; 2, 504 plaque samples over 11 mo. Cohort A: 16 children (8 healthy,	Single-tooth supragingival dental plaque (whole-mouth/site-resolved)	16S rRNA V3-V4 amplicon sequencing; Illumina MiSeq. A 24-sample subset underwent shotgun metagenomics on Illumina HiSeq 2500 for functional	Taxonomic ASVs, inferred functional pathways, and spatial clinical/microbial composite features.	Random-forest tooth-specific MiC; spatial-MiC (sMiC) adds surrounding-tooth clinical and microbiota features.	Cohort B internal 5-fold CV: local taxonomic MiC mean AUROC=0.89 for current caries and 0.82 for future onset; taxonomy-based sMiC	Independent in-study cohort A (same city/study, separate participants): direct sMiC validation AUROC=0.97

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						healthy, become carious, or progress.	8 caries-affected), 637 samples from all 20 teeth at 2 visits. Cohort B: 73 children, 1, 867 samples from 9 tooth sites over 5 visits. Shotgun functional subset: 24 samples from 7 cohort-B children.		calibration; PICRUSt2/Meta-Apo used for inferred functions.			AUROC=0.98 for current caries and 0.93 for future onset; function-based sMiC 0.90 and 0.81, respectively.	for disease-state classification and 0.96 for future caries-onset prediction. This is independent participant-cohort validation, not multicenter/geographic external validation.
108	Yuan, 2022	35859767	China (Shanghai)	Metabolic/endocrine: Type 1 diabetes	Disease detection; Phenotyping/stratification	New-onset T1D diagnosed by American Diabetes Association criteria vs healthy children attending routine checkups; chronic-phase T1D receiving insulin used for disease-phase comparison. Oral disease, active inflammatory/infectious disease, other severe organic/metabolic disease, and recent microbiota-altering treatment were excluded.	n=103 analyzed: 46 new-onset T1D, 10 chronic T1D, 47 healthy controls. Exact group age statistics: NR in the source report (Supplementary Table S1); prespecified age strata 3–7, >7–11, and >11 y. Chronic T1D insulin treatment duration averaged 2.57 y.	Unstimulated saliva	16S rRNA V3–V4 amplicon sequencing (338F/806R); Illumina MiSeq.	Taxonomic (genus-level OTUs) and community/diversity.	Genus-level random-forest classifiers based on discriminatory oral taxa; separate comparisons of new-onset T1D vs healthy and new-onset vs chronic treated T1D.	Same-cohort model assessment; resampling/split scheme NR. AUROC=0.933 for new-onset T1D vs healthy controls; AUROC=0.846 for new-onset vs chronic treated T1D.	No independent/external validation.
109	Alqaderi, 2026	41971337	Kuwait	Metabolic/endocrine: Type 1 diabetes	Disease detection	T1D participants recruited from diabetes centers/hospitals vs non-diabetic controls recruited from public schools across Kuwait; controls were age-matched. Participants with autoimmune diseases or cancer were excluded.	n=46 (23 T1D, 23 controls), eligibility age 10–21 y; actual age 17.22 ± 1.41 vs 17.78 ± 1.38 y. Mixed adolescent/young-adult cohort; not exclusively pediatric.	Paired saliva and supragingival dental plaque (Ramfjord teeth)	16S rRNA gene sequencing; QIIME 2 taxonomic processing with centered-log-ratio transformation.	Taxonomic (species-level) oral microbiome.	Random-forest classifiers for saliva and plaque, with and without clinical/demographic metadata.	Ten repeated random 75%/25% train/test partitions; hyperparameter tuning by 3-fold CV within training data. Saliva: accuracy ~94.2%, AUC ~0.94, sensitivity 96.7%, specificity 91.7% (with or without metadata). Plaque: accuracy ~73.3%, AUC ~0.75; with metadata sensitivity/specificity 73.3%/73.3%; microbiome-only 71.7%/75.0%.	No independent/external validation.
110	Yuan, 2022	36289225	China (nine regions: Harbin, Changchun, Taiyuan, Jinan, Zhengzhou, Suzhou, Shanghai,	Metabolic/endocrine: Type 1 diabetes	Disease detection	New-onset pediatric T1D diagnosed by American Diabetes Association criteria vs non-diabetic controls; discovery and	Discovery n=141 (64 T1D, 77 controls): age 7.53 ± 3.61 vs 7.92 ±	Stool	16S rRNA V3–V4 amplicon sequencing (Illumina MiSeq PE300); shotgun metagenomics;	Multi-omics: taxonomic species/genera, microbial functional features, and fecal	Random-forest models; genus model selected 22 key genera by leave-one-out CV. Primary	Discovery: genus-only AUC=0.815; 18-marker species+metabolite model	Independent in-study validation cohort: genus-only AUC=0.649;

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
			Nanchang, Fuzhou)			separately recruited validation cohorts were matched for age, sex, delivery, and feeding mode. DKA/non-DKA subgroup analyses were secondary.	2.92 y. Validation n=58 (29 T1D, 29 controls): age 6.58 ± 1.96 vs 6.48 ± 2.21 y.		fecal metabolomics by LC–MS/MS (ExionLC/AB Sciex QTRAP 6500+); SCFAs by GC–MS (Agilent 7890B/7000D).	metabolites/SCFAs.	multi-omics biomarker combined 9 bacterial species + 9 fecal metabolites.	AUC=0.976.	locked 9-species+9-metabolite panel AUC=0.809. Validation was separately recruited within the same multicenter study network, not an external publication/cohort system.
111	Liu, 2021	34040330	China (Linyi, Shandong; Hangzhou, Zhejiang)	Metabolic/endocrine: Type 1 diabetes	Disease detection	Children with newly confirmed T1DM vs age-, sex-, and education-matched healthy controls; T1DM diagnosed clinically and treated only with insulin. Recent microbiota-altering agents, active infection, IBS/IBD, and other autoimmune diseases were excluded.	n=98 (51 T1DM, 47 controls), age 6–14 y; mean age 10.38 ± 3.59 vs 9.58 ± 4.35 y.	Stool	16S rRNA V3–V4 amplicon sequencing; Illumina MiSeq.	Taxonomic bacterial microbiome with predicted functional profiles.	Four-taxon multivariable stepwise logistic-regression signature: Bacteroides vulgatus ATCC8482, B. ovatus, Eubacterium hallii group, and Anaerostipes hadrus; single-taxon ROC analyses also reported.	Same-cohort apparent ROC; no held-out/CV procedure reported. Single-taxon AUCs 0.294–0.690; four-taxon combination AUC=0.830.	Not evaluated: no independent/external cohort; authors state larger clinical validation is required.
112	Yuan, 2024	39027370	China (nine regions: Harbin, Changchun, Taiyuan, Jinan, Zhengzhou, Suzhou, Shanghai, Nanchang, Fuzhou)	Metabolic/endocrine: Type 1 diabetes	Disease detection	New-onset T1D children vs age- and sex-matched non-diabetic controls in a multicenter cross-sectional study. This appears to reuse the same 64 T1D/77-control parent cohort analyzed for gut multi-omics by Yuan et al. 2022 (PMID 36289225; No. 110), now profiling the blood microbiome.	n=141 (64 new-onset T1D, 77 controls), age 3–14 y. Exact age statistics are in referenced Table S1 and not reproduced in the source report.	Peripheral blood (with secondary comparison to previously generated gut/oral microbiome datasets)	16S rRNA V3–V4 amplicon sequencing (338F/806R); Illumina NovaSeq 6000, paired-end 2×250 bp; PICRUSt functional prediction.	Taxonomic (genus-level) blood microbiome and predicted functional pathways.	Random-forest classifier using 36 genus-level blood microbial features.	Cross-validation reported, but exact resampling scheme is NR in the source report. AUC=0.82.	No independent/external validation.
113	Huang, 2023	37170084	China (Hainan)	Metabolic/endocrine: Central precocious puberty	Disease detection	Girls with central precocious puberty vs healthy controls at Hainan Women and Children's Medical Center. CPP criteria included secondary sexual characteristics before age 8 y, gonadal development, growth spurt, pubertal gonadotropins/positive LHRH provocation test, and bone age >1 y advanced.	n=150 (91 CPP, 59 controls). CPP eligibility: girls <10 y; exact group age summary NR in the source report.	Stool (microbiome and fecal metabolomics); blood/plasma (metabolomics)	Fecal 16S rRNA V4 sequencing (Illumina MiSeq); untargeted fecal and blood metabolomics by LC–MS.	Multi-omics: taxonomic gut microbiota plus fecal and blood metabolites; classifiers were built by individual data type.	Random forest + Boruta feature selection after random 70%/30% train/test split. Microbiome model: 24 representative species. Fecal positive-ion metabolite model: 52 features; other metabolite models also explored.	Microbiome RF AUC: training 1.00, test 0.932, pooled 0.98. Fecal positive-ion metabolite RF AUC: training 1.00, test 0.85, pooled 0.988. Feature selection/modeling occurred within one cohort; nesting of feature selection relative to split is not fully specified.	No independent/external validation.
114	Yan, 2024	39039462	China (Shanghai, Taizhou, Shenyang, Hangzhou/Xiaoshan)	Metabolic/endocrine: Idiopathic short stature	Disease detection	ISS children hospitalized at four pediatric centers vs normal-height community controls; all Han Chinese. ISS required height <−2 SD, normal GH stimulation (>10 ng/mL), normal birth size, and exclusion of nutritional, endocrine, skeletal, genetic, metabolic, and chronic systemic causes.	n=51 (26 ISS, 25 controls), age 3–14 y.	Stool	Shotgun metagenomic sequencing (strain-level) plus untargeted fecal metabolomics.	Multi-omics: strain-level taxonomic features and fecal metabolites.	Random-forest classifier with recursive feature elimination; final 13-feature microbiome/metabolite panel. Random 70%/30% train/test split; five repeats of 10-fold CV used for model performance estimation.	Results/Figure report test AUC=0.933 (95% CI 0.799–1.00). Discussion reports AUC=0.953 for the test set; internal numerical inconsistency, therefore retained as Unclear rather than reconciled.	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
115	Li, 2022	35757432	China (East China; Shanghai)	Metabolic/endocrine: Idiopathic short stature	Disease detection	Prepubescent ISS patients attending Ruijin Hospital vs healthy normal-height East China controls with similar dietary habits. ISS required height <−2 SD, normal GH stimulation, normal birth size, slow growth, and exclusion of organic/genetic causes.	n=32 (16 ISS, 16 controls), age 4–8 y; median age 6.4 vs 6.6 y.	Stool (microbiome and SCFAs)	16S rRNA V3–V4 amplicon sequencing (338F/806R); fecal SCFAs by GC–MS.	Taxonomic microbiota plus metabolites (short-chain fatty acids).	Taxon ROC models and binomial logistic regression combining Faecalibacterium + Parabacteroides; fecal SCFAs evaluated individually.	Same-cohort apparent ROC. Faecalibacterium AUC=0.83 (95% CI 0.68–0.99), sensitivity 0.94, specificity 0.78; Parabacteroides AUC=0.79 (0.63–0.95), sensitivity 0.89, specificity 0.68; combined taxa AUC=0.88 (0.76–1.00), sensitivity 0.94, specificity 0.81. Butyrate AUC=0.94 (0.86–0.99), sensitivity 0.94, specificity 0.88.	No independent/external validation.
116	Gao, 2023	37602197	China (Shenzhen)	Metabolic/endocrine: Obesity	Disease detection	Chinese adolescents recruited at Shenzhen Children's Hospital and classified by WHO BMI-for-age into normal weight, overweight, or obese; diabetes, recent antibiotics, long-term medication, diarrhea, and chronic constipation were excluded.	n=205: normal weight 65, overweight 53, obesity 87; age 11–16 y. Mean age: 13.91 ± 1.61, 13.26 ± 1.50, and 12.90 ± 1.41 y, respectively.	Stool	16S rRNA V3–V4 amplicon sequencing; Illumina MiSeq; PICRUSt functional prediction.	Taxonomic OTUs/community with inferred functional pathways.	Random-forest classifier using 56 BMI-associated OTUs to classify BMI categories/obese phenotype.	Half of the dataset was used for training and half for testing; paper describes AUC from cross-validation testing results but does not further specify resampling. Obese-phenotype AUC=0.96.	No independent/external validation.
117	Li, 2019	31849851	China (Shenzhen)	Metabolic/endocrine: Undernutrition	Disease detection	Children <3 y with moderate/severe undernutrition (WAZ <−2 SD vs WHO reference) vs healthy children <3 y who passed physical examinations; recent antibiotics/probiotics/PPIs and major allergic, hereditary, inflammatory, infectious, endocrine, or metabolic conditions were excluded.	Total n=126 (65 undernourished, 61 healthy). Stage I discovery n=105 (54 UN, 51 controls); Stage II validation n=21 (11 UN, 10 controls). Age <3 y.	Stool	Shotgun metagenomics; paired-end 150-bp Illumina HiSeq. Stool SCFAs measured by gas chromatography (TRACE 1300 GC/FID).	Taxonomic species and functional microbial genes; SCFAs used for complementary metabolic association.	Random-forest/Wilcoxon feature selection: 16 microbial gene markers and 4 Bacteroides species markers. The article text lists B. uniformis twice among the four species markers.	Stage I discovery/feature selection; exact discovery AUC not separately reported in the main text.	Independent in-study Stage II validation on 21 additional children: mean AUC=0.87 for the 16-gene panel and 0.91 for the 4-species panel (five ROC repeats shown). Same center/study framework.
118	Sun, 2025	40340772	China (Huantai/Zibo, Shandong)	Cardiovascular: Pediatric hypertension	Disease detection	Children from the second follow-up of the Huantai Childhood Cardiovascular Health Cohort: hypertension defined by elevated BP on three separate occasions vs randomly selected children with normal BP. Recent antibiotics, vomiting/diarrhea, and gastrointestinal disease were excluded.	n=102 (51 hypertension, 51 normal BP), age 10–14 y.	Stool	16S rRNA V3–V4 amplicon sequencing (338F/806R); Illumina MiSeq; PICRUSt2 functional inference.	Taxonomic genus-level microbiota plus clinical–microbial composite.	Random forest used to screen genera; final discriminative composite combined BMI with norank_f__Lachnospiraceae and Dorea.	Same-cohort ROC without reported train/test split or CV. BMI alone AUC=0.91 (95% CI 0.85–0.97); BMI + two genera AUC=0.95 (0.91–0.99), category NRI=0.49, category-free NRI=0.75, IDI=0.12 (P<0.05).	No independent/external validation.
119	Zhang, 2026	41590860	China (Kunming, Yunnan)	Cardiovascular: Congenital heart disease-associated PAH	Disease detection; Phenotyping/stratification	Children with left-to-right shunt CHD and PAH (C-PAH; PASP >35 mmHg and PVR >3 Wood units) vs CHD without PAH (C-NPAH) and healthy controls. The clinically relevant within-CHD comparison assesses current	n=47: 15 C-PAH, 16 C-NPAH, 16 controls. Age, median [IQR]: 0.83 [0.4–2.5], 3.0 [2.0–3.8], and 4.0	Bronchoalveolar lavage fluid (BALF)	16S rRNA V3–V4 amplicon sequencing (Illumina MiSeq); untargeted metabolomics by UHPLC–Q Exactive HF-X LC–MS.	Multi-omics: airway taxonomic microbiota plus BALF metabolites.	RF classifiers using indicator-value or LEfSe microbial markers; differential-metabolite RF analyses; candidate metabolites included N1-methylnicotinamide	Microbial RF AUC distributions were repeated/validated 1, 000 times; density curves show AUC >0.7 for indicator-based and >0.8 for LEfSe-based classifiers. Exact	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						PAH complication status.	[3.5–7.2] y.				and 2-piperidone.	single AUC values are not numerically reported. Metabolite RF performance is reported graphically; exact AUC unavailable.	
120	Liu, 2020	33178621	China (Tianjin)	Hematology/oncology: Acute lymphoblastic leukemia	Disease detection	Newly diagnosed, untreated pediatric ALL meeting morphologic/cytochemical/immunophenotypic criteria vs healthy children from the same region; recent chemotherapy/radiotherapy, antibiotics/probiotics/NSAIDs, IBS, and major chronic diseases were excluded.	n=81 (58 ALL, 23 controls). Table 1: age 7.2 y (range 3–15) vs 7.6 y (3–16); Methods also describe mean ± SD 7.2 ± 3.6 and 7.6 ± 3.6 y, creating an age-summary reporting inconsistency.	Stool	16S rRNA quantitative oligonucleotide microarray (4, 751 probes covering 1, 485 bacterial species); selected taxa confirmed by qPCR.	Taxonomic species-level bacterial abundance.	Random-forest classifier based on bacterial abundance; species-level model was best. Important species included Tatumella ptyseos, Leptothrix discophora, and Brevibacterium casei.	Leave-one-out cross-validation: species-level AUC=0.843 (95% CI 0.744–0.942).	No independent/external validation.
121	Bai, 2017	28855025	China (Jinan, Shandong)	Hematology/oncology: Acute lymphoblastic leukemia	Disease detection	Children with newly diagnosed ALL (MICM criteria), sampled before chemotherapy, vs healthy community controls; both case/control groups were stratified by recent antibiotic exposure to examine confounding.	n=63: 30 ALL, 33 controls. Subgroups: ALL+ATBx n=20 (mean age 5.5 y, range 1.3–12.4), ALL without ATBx n=10 (5.6 y, 1.0–12.7); control+ATBx n=16 (4.3 y, 1.0–7.0), control without ATBx n=17 (4.5 y, 1.5–7.0).	Stool	16S rRNA V3–V4 amplicon sequencing (338F/806R); Illumina MiSeq.	Taxonomic bacterial community; order/family-level candidate biomarkers.	Single-taxon ROC analyses for Bacteroidales and Enterococcaceae, stratified by antibiotic exposure; no multivariable fixed panel.	Same-cohort apparent ROC. Without antibiotics: Bacteroidales AUC=0.735; Enterococcaceae AUC=0.724. With antibiotics: Enterococcaceae AUC=0.694.	No independent/external validation.
122	Liu, 2021	33898360	China (Shanghai)	Infectious: Sepsis	Disease detection	PICU children with sepsis by pediatric consensus criteria vs healthy children attending hospital health checkups. Sepsis group included 10 with confirmed pathogen infection and 10 without; all septic patients had received antibiotics before stool sampling.	n=40: 20 sepsis, 20 healthy controls. Age: controls 6.35 ± 2.00 y; sepsis with confirmed pathogen 3.40 ± 1.43 y; sepsis without confirmed pathogen 3.80 ± 2.30 y.	Stool	16S rRNA V4 amplicon sequencing (515F/806R); Illumina HiSeq 2500 PE250.	Taxonomic OTU-level gut microbiota.	Random-forest classifier using 20 differentially abundant taxonomic features identified in the microbiome-wide association analysis.	Same cohort; train/test or cross-validation scheme NR/Unclear. AUC=0.8125.	No independent/external validation.
123	Guo, 2025	40606147	China (Zhengzhou, Henan)	Ophthalmic: Retinopathy of prematurity	Disease detection	Preterm infants with ROP vs non-ROP infants at the First Affiliated Hospital of Zhengzhou University. Feces were sampled at postnatal week 2 (T1) and week 4 (T2); ROP was	Overall n=59 (29 ROP, 30 non-ROP). T1: 13 ROP/18 non-ROP samples; T2:	Stool	16S rRNA V3–V4 amplicon sequencing (341F/805R), Illumina NextSeq 2000; T2 untargeted fecal metabolomics by Vanquish	Multi-omics: taxonomic gut microbiota plus fecal metabolites.	At T2, LEfSe differential genera and top-20 differential accumulated metabolites were evaluated by random-forest ROC; bacterial-only,	Same-cohort apparent ROC after feature discovery on the same T2 data: bacterial community AUC=0.9958; metabolite panel	No independent/external validation.

No.	First author, year	PMID	Country/region	Disease domain/indication	Clinical task	Population/comparator	Sample size/age	Matrix	Analytical platform	Biomarker layer	Biomarker/model	Development/internal performance	Independent/external validation
						generally diagnosed at week 4. All received cephalosporins; probiotic use differed between groups.	20 ROP/24 non-ROP samples. Sixteen infants provided samples at both times. Gestational age: 29.55 ± 1.75 vs 31.83 ± 1.52 weeks; birth weight 1162.94 ± 286.22 vs 1680.66 ± 443.08 g. ROP: stage 1 n=20, stage 2 n=9.		UHPLC–Q Exactive Focus LC–MS.		metabolite-only, and combined signatures were compared. No train/test/CV scheme reported.	AUC=1.00; combined AUC=1.00. These are not independent validation estimates.	
124	You, 2023	37007464	China (Changsha, Hunan)	Neonatal/perinatal: Pathologic neonatal jaundice	Differential diagnosis	Hospitalized neonates with jaundice: pathologic jaundice vs physiologic jaundice according to clinical diagnostic criteria. Sepsis, intracranial infection, and digestive-system disease were excluded; the comparator was not a healthy non-jaundiced group.	n=28 (14 pathologic jaundice, 14 physiologic jaundice). Age, median [IQR]: 10.50 [5.00-14.25] vs 12.00 [5.00-18.00] days; gestational age, mean ± SD: 36.95 ± 2.45 vs 35.51 ± 3.02 weeks.	Stool; first defecation after hospitalization	16S rDNA V4-V5 amplicon sequencing (515F/907R); Illumina NovaSeq 6000; QIIME2/97% OTU-based analysis.	Taxonomic and community/diversity features.	Single-taxon ROC candidates; Bacteroidetes was the primary marker. Additional candidates: Actinobacteria, Bifidobacterium, Rothia, Planctomycetes, and Bradyrhizobium. No fixed multivariable classifier.	Apparent same-cohort ROC analysis; no cross-validation or held-out test set reported. Bacteroidetes AUC=0.839 (95% CI 0.648-0.995); Actinobacteria 0.827; Bifidobacterium 0.763; Rothia 0.724; Planctomycetes 0.755; Bradyrhizobium 0.763.	No independent/external validation. Antibiotic exposure was common in both groups, and the authors noted residual infection/antibiotic confounding and lack of longitudinal follow-up.

Note: Clinical-task labels were standardized across records. Performance metrics are reproduced only when explicitly reported in the source publications; no unreported values were inferred or calculated. AUC, AUROC, and ROC-AUC terminology is retained as reported by the source studies. “Biomarker layer” denotes the biological feature layer (e.g., taxonomic/community, functional, genomic-variant, metabolite/VOC, or multi-omics features), whereas “Biomarker/model” denotes the operational biomarker set and/or classifier. Each record is indexed by its PubMed identifier (PMID), which can be entered directly into PubMed to retrieve the bibliographic record and links to available full-text sources. Study-specific cohort names, software packages, assay names, and commercial platforms are retained as reported in the source publications.

Abbreviations: AD, atopic dermatitis; ADHD, attention-deficit/hyperactivity disorder; AP, acute pancreatitis; ASD, autism spectrum disorder; ASV, amplicon sequence variant; AUC, area under the curve; AUROC, area under the receiver operating characteristic curve; BALF, bronchoalveolar lavage fluid; BCFA, branched-chain fatty acid; BMI, body mass index; CART, classification and regression tree; CD, Crohn disease; CeD, celiac disease; CI, confidence interval; CMPA, cow’s milk protein allergy; CRP, C-reactive protein; CV, cross-validation; ECC, early childhood caries; EEN, exclusive enteral nutrition; FDR, false discovery rate; FI, feeding intolerance; FT, feeding tolerance; GC-IMS, gas chromatography–ion mobility spectrometry; GC-MS, gas chromatography–mass spectrometry; GFD, gluten-free diet; GI, gastrointestinal; HC, healthy control(s); HILIC, hydrophilic interaction liquid chromatography; HPLC, high-performance liquid chromatography; IBD, inflammatory bowel disease; IBD-U, inflammatory bowel disease unclassified; IBS, irritable bowel syndrome; InDel, insertion/deletion; IQR, interquartile range; ITS, internal transcribed spacer; JIA, juvenile idiopathic arthritis; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG ortholog; LASSO, least absolute shrinkage and selection operator; LC-MS, liquid chromatography–mass spectrometry; LOOCV, leave-one-out cross-validation; LRTI, lower respiratory tract infection; MAG, metagenome-assembled genome; MASLD, metabolic dysfunction-associated steatotic liver disease; MCC, Matthews correlation coefficient; MDD, major depressive disorder; ML, machine learning; MPP, *Mycoplasma pneumoniae* pneumonia; NA, not applicable; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; NMR, nuclear magnetic resonance; NPV, negative predictive value; NR, not reported; NT, neurotypical; OOB, out-of-bag; OSA, obstructive sleep apnea; OTU, operational taxonomic unit; PCR, polymerase chain reaction; PLS-DA, partial least-squares discriminant analysis; PMID, PubMed identifier; PPV, positive predictive value; PR-AUC, area under the precision–recall curve; PSC, primary sclerosing cholangitis; RF, random forest; ROC, receiver operating characteristic; ROC-AUC, area under the receiver operating characteristic curve; ROP, retinopathy of prematurity; RRTI, recurrent respiratory tract infection; SCFA, short-chain fatty acid; SD, standard deviation; SEM, standard error of the mean; SNV, single-nucleotide variant; SV, structural variant; SVM, support vector machine; T1D/T1DM, type 1 diabetes/type 1 diabetes mellitus; TD, typically developing; UC, ulcerative colitis; UHPLC, ultra-high-performance liquid chromatography; URTI, upper respiratory tract infection; VLBW, very low birth weight; VOC, volatile organic compound; WGS, whole-genome sequencing.