Received: 30 August 2023 | Revised: 5 March 2024 | Accepted: 7 March 2024

DOI: 10.1111/cea.14475

ORIGINAL ARTICLE

WILEY

An integrated molecular risk score early in life for subsequent childhood asthma risk

Andreas Böck[1,2] | Kathrin Urner[1,2] | Jana Kristin Eckert[1,2] | Michael Salvermoser[1,2] | Kristina Laubhahn[1,3] | Sonja Kunze[4,5] | Jörg Kumbrink[6] | Marc P. Hoeppner[7] | Kathrin Kalkbrenner[1,2] | Simone Kreimeier[2,8] | Kirsten Beyer[2,9] | Eckard Hamelmann[2,10] | Michael Kabesch[2,11] | Martin Depner[2,12] | Gesine Hansen[2,13,14,15] | Josef Riedler[16] | Marjut Roponen[17] | Elisabeth Schmausser-Hechfellner[12] | Cindy Barnig[18,19] | Amandine Divaret-Chauveau[20,21,22] | Anne M. Karvonen[23] | Juha Pekkanen[23,24] | Remo Frei[25,26] | Caroline Roduit[25,26,27,28] | Roger Lauener[25,27] | Bianca Schaub[1,2,3] | CHAMP study group; PASTURE study group

Correspondence
Bianca Schaub, Department of Allergy/Immunology, University Children's Hospital, Dr. von Haunersches Kinderspital, Lindwurmstraße 4, 80337 Munich, Germany.
Email: bianca.schaub@med.uni-muenchen.de

Funding information
Juho Vainion Säätiö; Reasearch Committee of the Wellbeing Services County of North Savo (Kuopio University Hospital) for the State Research Funding (EVO/VTR); Farmers' Social Insurance Institution (Mela); Finnish Cultural Foundation; Yrjö Jahnssonin Säätiö; Kuopio Area Respiratory Foundation (AMK); Kühne Foundation; Finnish Institute for Health and Welfare from the Academy of Finland, Grant/Award Number: 139021, 287675, 296814, 296817 and 308254; Deutsche Forschungsgemeinschaft, Grant/Award Number: SCHA 997/11-1; Bundesministerium für Bildung und Forschung, Grant/Award Number: 01GL1742A, 01GL1742B, 01GL1742C, 01GL1742D, 01GL1742E and 01GL1742F; Hauner Verein

Abstract

Background: Numerous children present with early wheeze symptoms, yet solely a subgroup develops childhood asthma. Early identification of children at risk is key for clinical monitoring, timely patient-tailored treatment, and preventing chronic, severe sequelae. For early prediction of childhood asthma, we aimed to define an integrated risk score combining established risk factors with genome-wide molecular markers at birth, complemented by subsequent clinical symptoms/diagnoses (wheezing, atopic dermatitis, food allergy).

Methods: Three longitudinal birth cohorts (PAULINA/PAULCHEN, $n = 190 + 93 = 283$, PASTURE, $n = 1133$) were used to predict childhood asthma (age 5–11) including epidemiological characteristics and molecular markers: genotype, DNA methylation and mRNA expression (RNASeq/NanoString). Apparent (ap) and optimism-corrected (oc) performance (AUC/R2) was assessed leveraging evidence from independent studies (Naïve-Bayes approach) combined with high-dimensional logistic regression models (LASSO).

Results: Asthma prediction with epidemiological characteristics at birth (maternal asthma, sex, farm environment) yielded an ocAUC = 0.65. Inclusion of molecular markers as predictors resulted in an improvement in apparent prediction performance, however, for optimism-corrected performance only a moderate increase was

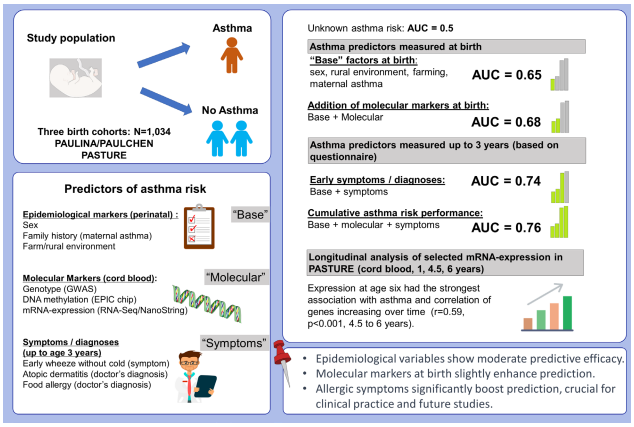
CHAMP study group and PASTURE study group members are in the Appendix.

For affiliations refer to page 325.

observed (upto ocAUC = 0.68). The greatest discriminate power was reached by adding the first symptoms/diagnosis (up to ocAUC = 0.76; increase of 0.08, $p = .002$). Longitudinal analysis of selected mRNA expression in PASTURE (cord blood, 1, 4.5, 6 years) showed that expression at age six had the strongest association with asthma and correlation of genes getting larger over time ($r = .59$, $p < .001$, 4.5–6 years).

Conclusion: Applying epidemiological predictors alone showed moderate predictive abilities. Molecular markers from birth modestly improved prediction. Allergic symptoms/diagnoses enhanced the power of prediction, which is important for clinical practice and for the design of future studies with molecular markers.

KEYWORDS
asthma, epidemiology, genetics, paediatrics, prevention



GRAPHICAL ABSTRACT

For early prediction of childhood asthma, an integrated risk score combining established risk-factors with genome-wide molecular markers (genotype, DNA methylation, mRNA expression) at birth was defined, complemented by subsequent clinical symptoms/diagnoses (wheeze, atopic dermatitis, food allergy). While epidemiological predictors have moderate predictive power, molecular markers from birth improve prediction to a modest extent. Including allergic symptoms/diagnoses significantly increased predictive efficiency.

1 | INTRODUCTION

Asthma is one of the most common chronic lung diseases in childhood with varying prevalence across Europe, partly still increasing.[1–3] Around one-third of preschool children aged 1–5 years have asthma-like symptoms such as wheezing and 30–40% of them will develop asthma in later childhood.[4,5] Yet, effective and clinically applicable tools for early identification of children at asthma risk are lacking. Although numerous independent asthma prediction models were proposed,[5–9] they showed poor predictive accuracy when externally validated[10] and application is not recommended in guidelines due to limited performance for clinical practice (GINA).[11] These models typically rely on current symptoms like early wheezing or allergic sensitization.[12] Before occurrence of symptoms, cord blood sampling allows non-invasive extraction of biomarkers, including genetic variants and DNA methylation. Thus, numerous variants influencing asthma susceptibility identified through genome-wide association studies[13–16] can be measured already at birth. Epigenetic mechanisms add to asthma risk, with DNA methylation potentially playing an important role in asthma development.[17,18] However, neither genetic nor epigenetic biomarkers alone could accurately predict the development of asthma at an early age.[19] Although transcriptomic studies uncovered underlying mechanisms of asthma[20–24] at manifestation, to our knowledge, no asthma predictive transcription markers for this time window were identified yet.

Key messages

- Epidemiological predictors exhibit moderate predictive efficacy for childhood asthma.
- Molecular markers from birth modestly enhance prediction.
- Inclusion of allergic symptoms/diagnoses significantly boosts predictive power, informing future study designs.

We aimed to develop a novel prediction score by simultaneously integrating genotype, DNA methylation and mRNA expression data derived from cord blood on top of well-recognized and clinically easy-to-obtain epidemiological factors. Considering generalizability and transferability of prediction to independent populations, we based the predictor selection whenever possible on evidence from prior independent studies. We extended our score by including the first symptoms/diagnoses of wheezing, atopic dermatitis (AD) or food allergy (FA) occurring until age 3 years.

2 | METHODS

2.1 | Study population

For PAULINA and PAULCHEN ($n = 190/93$), pregnant women from urban (Munich) and rural areas were recruited from 2004 to 2009 and $N = 229$ included based on strict inclusion criteria.[25,26] ISAAC-based questionnaires were completed by parents after birth and at follow-ups (FU) after 3, 6 and 10 years (FU3/FU6/FU10). PASTURE is a prospective birth cohort from rural areas in five European countries including 1133 families recruited from 2002 to 2005 (530 farming, 603 non-farming), with 10 FU up to age 16 years.[27]

For all studies, asthma was defined identical as a reported doctor's diagnosis of asthma or recurrent obstructive-, or asthmatic bronchitis after age 5 years. If the reported doctor's diagnosis was before age 5 years, additional asthma symptoms and/or medication after age 5 years had to be reported (Section S3.1).

2.2 | Study participants with biomaterial and derived data

Selection of subsets for molecular measurements included all children from PAULINA/PAULCHEN and a nested case–control design for the PASTURE cohort. Samples of all asthma cases meeting quality requirements were chosen. Non-asthma controls were selected to maximize overlap in DNA methylation and RNAseq data, considering the status of AD, sex, farming and study centre.

Cord blood was sampled for genotyping (GSA-chip), DNA methylation (Illumina-EPIC) and mRNA expression analyses (NanoString for PAULINA/PAULCHEN, RNAseq for PASTURE) (Figure 1). Additionally, PASTURE blood samples were taken at 1, 4.5 and 6 years for qPCR analysis (Table S1, Sections S3 and S4).

2.3 | Established risk factors

To enhance applicability and limit statistical complexity due to high-dimensional variable selection, an evidence-based selection of variables of well-replicated literature-based results was performed.

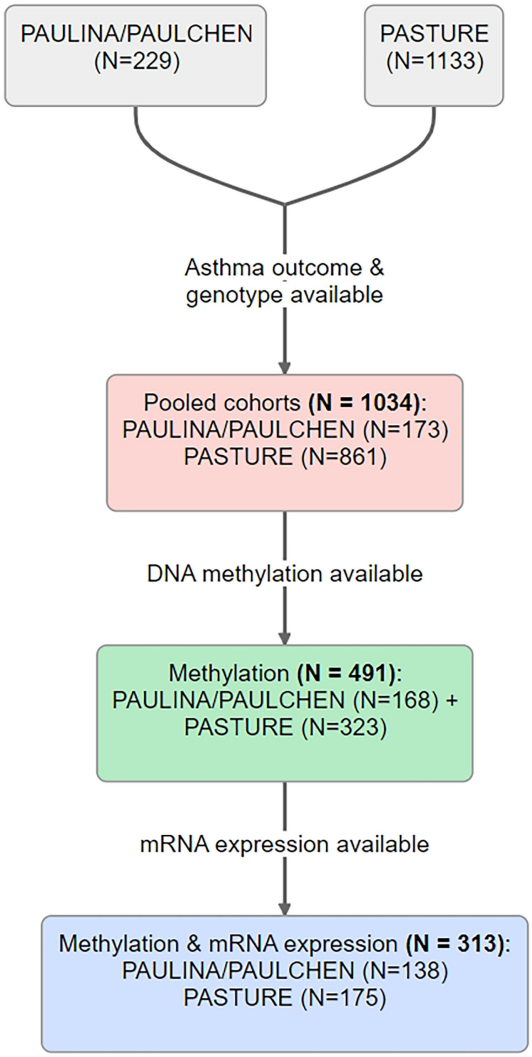


FIGURE 1 Flowchart sample size. Subsets of data with respective outcome and molecular data available (details in Section 2).

2.3.1 | Base variables

Previously identified as asthma risk-associated variables sex,[28] maternal asthma[29] and farming and rural status[30–35] served as *base variables* being included in every prediction model.

2.3.2 | Genetic variant score (variant score)

Different published variant scores were evaluated, mostly leading to comparable results (Tables S2 and S3). To incorporate genetic predisposition based on genotype, we utilized the largest meta-analysed association study on childhood asthma,[14] resulting in 12 genetic variants incorporated into the *variant score* (Supplement S3.8.1, Table S4).

2.3.3 | Methylation score (methyl. score)

Similarly, based on an epigenome-wide association study (EWAS) of cord blood DNA and childhood asthma,[18] eight differentially methylated CpGs were integrated into the *methylation score* using the reported OR (Table S5, Supplement S3.8.2).

2.3.4 | Diagnoses/symptoms (diag/symp)

Three early diagnoses/symptoms were included. Early wheezing without cold[5,8,9] was defined as any wheezing symptoms outside cold episodes, early FA[36] and AD[5,8,9,37] as reported doctor's diagnoses, all in the first 3 years of life.

2.4 | Priorization of variables in high-dimensional data sets

2.4.1 | Prioritized DNA methylation markers

EPIC-DNA methylation sites comprised 666,727 predictor variables after QC. From these, we selected 162 *prioritized sites* located within 34 differentially methylated regions (Table S6).[18]

2.4.2 | Prioritized mRNA expression markers

The entire mRNA expression data comprised 14,492 variables. For mRNA, no transcriptome-wide association study linking expression in cord blood to childhood asthma is available to our knowledge. However, we defined a *prioritized set* of 167 genes by reviewing relevant literature reporting associations of asthma and mRNA expression from later time points during childhood beyond cord blood (Table S7).[20–24,38–41]

2.5 | Statistics

Variables were described with (relative) frequencies, ORs and descriptive *p*-values. Residual heterogeneity was assessed with likelihood ratio tests between models with/without a main and interaction effect of study (Section S3.7).

Logistic regression models were used for predictive modelling. Discrimination was reported as area under the receiver-operating characteristics curve (AUC), however, Nagelkerke's pseudo-R^2,[42] which in addition considers the calibration aspect, was used as the main criterion. Models were internally validated by bootstrapping which was suggested as a preferable method compared to a split-sample approach.[43] Performance was presented as *apparent (ap)*, when entire data were used for model fitting and evaluation, and *optimism corrected (oc)*, when the in-bag bootstrap samples were used in model fitting and the out-of-bag samples for evaluation. Model and variable selection was based on optimal *oc* performance. Further details, including the logistic regression models are provided in the supplement (Section S3.7, Figures S1–S4).

2.6 | Sensitivity analyses

Sensitivity analyses comprised a doctor's diagnosis of asthma solely as outcome definition, an assessment of several variant scores, and different pre-processing for methylation data (Section S4).

3 | RESULTS

3.1 | Characterization of study cohorts

Children from three longitudinal birth cohorts ($N=1034$) were included based on available asthma status and GWAS data (173 from PAULINA/PAULCHEN, 861 from PASTURE). Of these, 491 children had DNA methylation data ($n=168$ PAULINA/PAULCHEN; $n=323$ PASTURE) and for 353 children mRNA expression profiles were analysed ($n=141$ PAULINA/PAULCHEN; $n=212$ PASTURE). The subset of $n=313$ children had all three molecular components available (genotype, DNA methylation, mRNA expression, Figure S1). This resulted in data subsets enriched for asthma and not representative of the whole cohort (Tables S8 and S9).

3.2 | Associations and predictive performance of established risk factors

As base components, we assessed the association of established risk factors for childhood asthma, sex, maternal asthma and rural/farming status for compatibility with the literature (Table 1). In the pooled samples ($n=1034$), it was more likely for males (OR = 2.00) and children with asthmatic mothers (OR = 3.63) to develop childhood asthma. Farm environment (OR = 0.65) and rural areas (OR = 0.88) had protective effects. Heterogeneity between the cohorts was observed in unadjusted analyses when the individual risk factors/covariates were considered separately. Controlling for sex, maternal asthma and rural and farming environment as given by the literature, heterogeneity was reduced to a statistically non-significant level, which made it possible to analyse the cohorts together. Associations of genetic variants included in the variant score with childhood asthma risk were detectable in our cohorts (*p*-values .001–.085, Table S4). Effect sizes of DNA methylation included in the methylation score were mostly of similar magnitude as reported, yet without statistical significant associations in our cohorts, Table S5. Early symptoms and atopic diseases were positively associated with asthma: early wheeze without cold OR = 4.15, OR of AD = 2.22, OR of FA = 4.64 (Table 1).

Thus, observing compatible effect sizes, we integrated the base components using the Naïve-Bayes approach as it led to higher optimism-corrected performance in all three data subsets compared to maximum likelihood estimation (Table S10). Their combined

TABLE 1 Odds ratios (OR): Associations with childhood asthma.

Term	PAULINA/PAULCHEN $N = 173$, 29 (16.8%) asthmatics		PASTURE $N = 861$, 92 (10.7%) asthmatics		Pooled cohorts*	Naïve-Bayes-plug-in#	Heterogeneity between cohorts (p-value)	
	Prevalence of term (%)	Odds ratio for asthma (95% CI)	Prevalence of term (%)	Odds ratio for asthma (95% CI)	Odds ratio for asthma (95% CI)	Odds ratio for asthma	Unadjusted†	Covariate adjusted‡
Maternal Asthma	9.25	2.52 (0.74–7.62)	8.94	4.08 (2.31–7.01)	3.63 (2.17–5.95)	3.0[a]	.07	.50
Male sex	51.45	1.68 (0.75–3.92)	51.45	2.10 (1.34–3.37)	2.00 (1.34–3.02)	1.6[b]	.08	.54
Farming environment	10.98	1.93 (0.58–5.60)	48.66	0.58 (0.37–0.91)	0.65 (0.42–0.99)	0.60[c]	.06	.07
Rural vs urban Environment	41.62	1.38 (0.62–3.10)	100	–	0.88 (0.48–1.68)	0.83[d]	–	–
Wheeze without Cold in the first 3 years of life	7.51	3.54 (1.00–11.56)	14.87	5.01 (3.12–8.01)	4.15 (2.64–6.49)	5.0[e]	.01	.18
AD diagnosis in the first 3 years of life	14.45	1.29 (0.40–3.56)	21.60	2.80 (1.77–4.39)	2.22 (1.44–3.38)	2.1[f]	.02	.19
FA diagnosis in the first 3 years of life	2.31	16.50 (2.03–340.68)	6.74	4.87 (2.63–8.78)	4.64 (2.54–8.29)	2.5[g]	.02	.20

Note: Odds ratio (OR) and 95% confidence interval (95% CI).

Abbreviations: AD, atopic dermatitis; FA, food allergy.

*Reported effects in the pooled cohort are mutually adjusted for maternal asthma, sex, farming- and rural environment #OR from literature were recapped and used as-is in the Naive Bayes approach (plug-in). Used literature was [a]OR = 3.04,[29] [b]OR = 1.45 (Avon Longitudinal Study of Parents and Children), OR = 1.42 (Millennium Cohort Study)[28]; OR = 1.6–1.7,[62] [c]OR = 0.77[30]; OR = 0.51[31]; OR = 0.57–0.90[32]; OR = 0.63[33]; OR = 0.62,[34] [d]OR = 0.83,[35] [e]OR = 4.8[9]; OR = 6[5]; OR = 2.9,[8] [f]OR = 1.9[9]; OR = 2.3[5]; OR = 2.1[8]; OR = 2.8,[37] [g]OR = 2.4.[36]

†Likelihood ratio test for heterogeneity between cohorts, unadjusted for further covariates.

‡Test for heterogeneity between cohorts, adjusted for maternal asthma, sex, farming and rural environment as reported in the literature (Naïve-Bayes-plug-in as offset term).

TABLE 2 Variant and methylation score effect.

Data subset	Model	AUC (ap/oc)	R2 (%) (ap/oc)	Increase AUC vs. base (oc)	Increase R2 (%) vs. base (oc)	Likelihood ratio test vs. base (*p*)
1034	Base	0.651/0.652	7.0/6.1			
	+variant score	0.681/0.677	8.9/7.5	0.025	1.4	.001
491	Base	0.595/0.596	3.8/2.2			
	+variant score	0.627/0.621	5.2/2.3	0.025	0.1	.038
	+methyl. score	0.631/0.625	5.0/2.5	0.029	0.3	.051
	+variant and methyl. score	0.648/0.635	6.4/2.7	0.039	0.5	.015 .043[a]
313	Base	0.607/0.610	5.4/2.9			
	+variant score	0.630/0.626	7.2/3.3	0.016	0.4	.058
	+methyl. score	0.628/0.614	5.9/2.0	0.005	No increase	.339
	+variant and methyl. score	0.637/0.625	7.8/2.6	0.016	No increase	.094 .286[a]

Note: Mean performance (500 bootstrap replications) when adding the variant and/or the methylation score to the model including base variables (maternal asthma, sex, farming and rural status).

Abbreviations: ap, apparent; AUC, area under the curve; oc, optimism corrected; R2, Proportion of explained variation (Nagelkerke's measure of determination).

[a]Test against model including base + variant score.

apparent (optimism-corrected) performances were apAUC=0.651 (ocAUC=0.652) and apR2=7.0% (ocR2=6.1%) for the whole data-set $n=1034$. In the $n=313$-subset, the base components yielded an apAUC=0.607 (ocAUC=0.610) and apR2=5.4% (ocR2=2.9%) (Table 2, base model).

3.3 | Inclusion of established molecular components to prediction

By integrating the genetic variant score into the base prediction model, the apAUC and ocAUC improved marginally by 0.030/0.025 ($n=1034$), while the methylation score increased apAUC by 0.036 (ocAUC by 0.029; $n=491$) (Table 2). The combination of both enhanced apAUC by 0.054 (ocAUC by 0.039; $n=491$). Lower predictive improvements were observed in the subset with complete molecular data ($n=313$).

3.4 | Data-driven selection of additional molecular markers to improve prediction

Next, the high-dimensional data sets (mRNA and methylation data) were assessed for predictive utility in different combinations of predictor setups (Table S11): (i) prioritized mRNA expressions (187 variables), (ii) all mRNA expressions (14,492 variables), (iii) prioritized DNA methylation sites (162 variables) and (iv) all DNA methylation sites (666,727 variables).

In the data set $n=313$, the highest performance was reached by selecting variables from the entire set of mRNA expression data (model_7, Figure 2a). The combined effect of the four genes AIDA, METTL4, ZDHHC2 (higher expression associated with higher asthma risk) and RNF25 (overexpression contributing to lower asthma risk) (Table S12a) increased the apAUC (model_3 to 7) from 0.64 to 0.69 (ocAUC 0.63–0.64). With focus on optimism-corrected performance, the addition of the entire methylation data (model_8) or the entire mRNA expression and methylation data in combination (model_9) showed no impact on predictive performance and similar low effects. The prioritization of markers (model_4/5/6), although statistically beneficial by reducing dimensionality, did not prove helpful. All the above models could not improve on the simpler model 3 with an ocAUC of 0.63.

Additionally, the methylation data were assessed in the larger sub-sample ($n=491$) where no complete mRNA measurements were available. Here, highest predictive performance was observed selecting three additional CpG sites (Table S12b) from the entire methylation set enhancing apAUC (model_3 to 8) from 0.65 to 0.72 (ocAUC=0.63–0.67). In contrast, with prioritized methylation data (model_4), no improvement in performance to simpler model 3 was observed apAUC=0.65 (ocAUC=0.63) (Figure 2a). No additional molecular data were available for the $n=1034$ data set.

3.5 | Considering early symptoms/diagnoses

As common in prediction scores for asthma, we also considered early diagnoses and symptoms and extended the model by FA, AD and early wheeze without cold. Starting with base variables (Figure 2a; $n=1034$, model_1, apAUC/ocAUC=0.65) the inclusion of symptoms/diagnoses ($n=1034$, model_1) improved the AUC on average by 0.09 to apAUC/ocAUC=0.74 (Table 3, all $p<.01$).

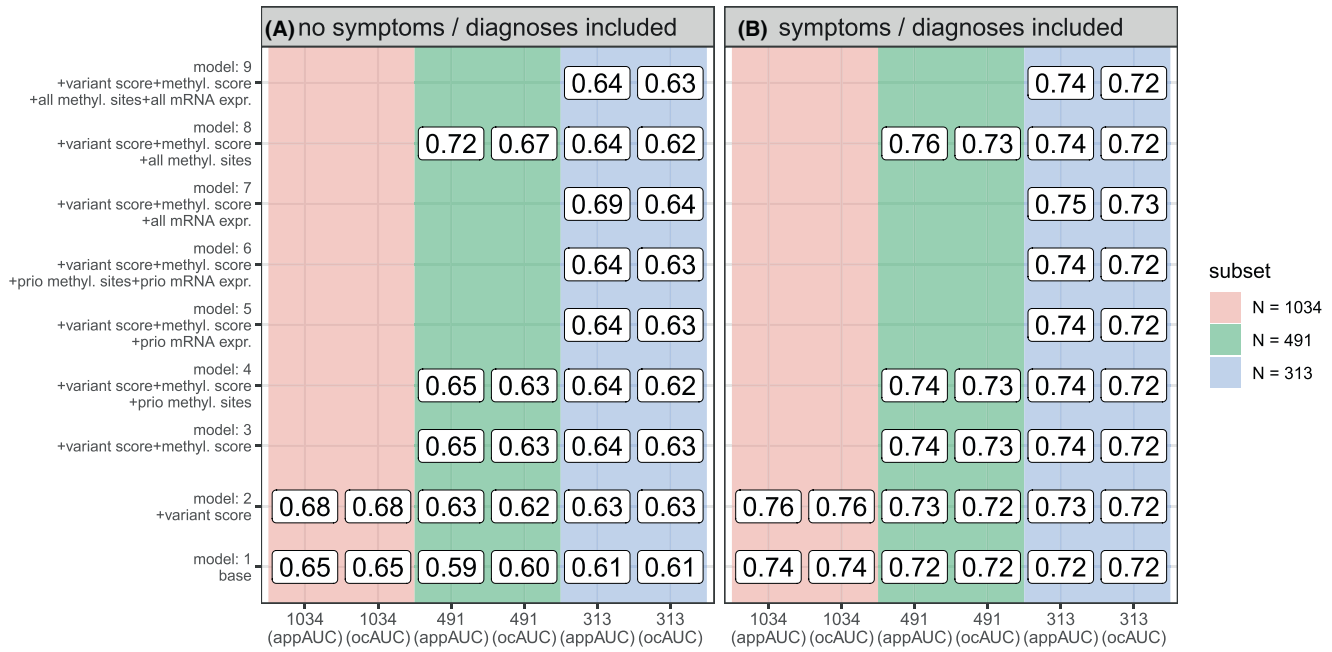


FIGURE 2 Predictive model performance. Apparent (ap) and optimism-corrected (oc) AUC is shown for models with data-driven selection of additional molecular markers (models 4–9) in contrast to respective models without newly selected molecular components (models 1–3). Models are sorted and numbered by model complexity (simple models at bottom) and grouped into models without (A) or with symptoms/diagnoses included (B). 'base' model includes variables sex, maternal asthma, farming status and rural area status; base variables are part of every model. Variant score is built from 12 Chromosome 17 SNPs, Methylation score is built from 8 methylation sites. Symptoms/diagnoses include early food allergy, early eczema and early wheeze without cold. 'all mRNA expr.' include 14,492 mRNA expression variables, prio a subset of 154. 'all methyl. sites' include 666,727 DNA methylation variables, prio a subset of 162.

Considering the high-dimensional molecular data in combination with symptoms/diagnoses, we used an analogue setup as above. With symptoms/diagnoses included (Figure 2b), the highest predictive performance in data subset $n=313$ was reached by selecting variables from the entire set of mRNA expression data (model_7). The combined effect of three selected genes (AIDA, ZDHHC2-overexpression contributing to higher asthma risk, and TREX1-overexpression contributing to lower asthma risk) increased the apAUC (model_3 to 7) from 0.74 to 0.75 (ocAUC = 0.72–0.73) (Figure 2, Table S12c). With focus on oc performance, neither the entire methylation data (model_8) or the entire mRNA expression and methylation data in combination (model_9) nor the prioritized marker sets (model_4, 5, 6) did substantially add to predictive performance, all showing an ocAUC of 0.72 as the simpler model 3.

Again, methylation without mRNA expression data was assessed in the larger subset ($n=491$). By selecting 3 additional CpG sites (Table S12d) from the entire methylation set (model_8), apAUC could be enhanced from 0.74 to 0.76 (model_3 to 8; ocAUC from 0.73 to 0.73). Two CpGs were in overlap with the model without symptoms (Table S12b). The approach restricted to prioritized methylation data (model_4) showed no improvement on simpler model 3.

In summary, significant improvement and highest absolute prediction performance were observed when including symptoms/diagnoses (Table 3, on average + 0.09 in ocAUC, and + 11.9% in ocR2). Nevertheless, single new CpG and mRNA expression markers were identified as contributing to risk prediction, albeit to a modest absolute scale.

3.6 | Association of mRNA expression to asthma increases over time until age six

As we observed low predictive performance with biomarkers at birth, we studied the temporal association between asthma and gene expression also at later time points of life (1, 4.5 and 6 years) in depth. Thus, in 40 genes measured over multiple time points in the PASTURE cohort (qPCR, cord blood, 1 year, 4.5 years, 6 years, Table S13), the association of the same gene-set from several time points to a uniformly defined asthma outcome was inferred.

Including covariates sex, maternal asthma, farming and genetic variant score, increased performance was solely observed at 6 years (apAUC = 0.71–0.79; ocAUC = 0.68–0.72) by combining mRNA expression of genes SregionIgE, SOCS2 (overexpression higher asthma risk) and IL10, TNFSF13B, LY96 (overexpression lower asthma risk) (Table 4).

The temporal correlation of mRNA expression was assessed between each time point to one another (six coefficients) separately by gene. Overall average correlation between all time points and genes was mean_$r=.08$ (Figure 3). Between years 4 and 6 correlation is twice as high ($r=.16$), compared to the correlation of cord blood and year 6 three times higher. SregionIgE showed highest temporal correlation ($r=.59$, $p<.001$, between year 4 and 6; $r=.45$ between year 1 and year 4, $p<.001$) (Figure 3; Table S14).

Sensitivity analyses revealed large impact of pre-processing of methylation data, and lower predictive performances when

TABLE 3 Contribution of symptoms/diagnoses.

		No symptoms/diagnoses included				Symptoms/diagnoses included				Increase				
		AUC		R2 (%)		AUC		R2 (%)		AUC		R2 (%)		
Model	*N*	ap	oc	ap	oc	ap	oc	ap	oc	ap	oc	ap	oc	*p*-Value[a]
Model 1: base	1034	0.65	0.65	7.0	6.1	0.74	0.74	16.8	16.2	0.09	0.09	9.8	10.1	.001
	491	0.59	0.60	3.8	2.2	0.72	0.72	18.1	16.6	0.12	0.12	14.3	14.4	<.001
	313	0.61	0.61	5.4	2.9	0.72	0.72	18.8	16.7	0.11	0.11	13.3	13.8	.004
Model 2: +variant score	1034	0.68	0.68	8.9	7.5	0.76	0.76	18.2	17.1	0.08	0.08	9.3	9.6	.002
	491	0.63	0.62	5.2	2.3	0.73	0.72	18.5	15.9	0.10	0.10	13.4	13.6	.001
	313	0.63	0.63	7.2	3.3	0.73	0.72	19.1	15.6	0.10	0.10	11.9	12.3	.007
Model 3: +variant +methyl. score	491	0.65	0.63	6.4	2.7	0.74	0.73	19.1	15.7	0.09	0.09	12.7	13.0	.001
	313	0.64	0.63	7.8	2.6	0.74	0.72	19.3	14.3	0.10	0.10	11.5	11.7	.005
Model 4: +variant +methyl. score +prio methyl. sites	491	0.65	0.63	6.4	2.4	0.74	0.73	19.1	15.6	0.09	0.09	12.7	13.2	.001
	313	0.64	0.62	7.8	2.2	0.74	0.72	19.3	13.8	0.10	0.09	11.5	11.6	.005
Model 5: +variant+methyl. score + prio mRNA expr.	313	0.64	0.63	7.8	2.3	0.74	0.72	19.3	14.2	0.10	0.10	11.5	11.9	.007
Model 6: +variant +methyl. score + prio methyl. sites +prio mRNA expr.	313	0.64	0.63	7.8	2.3	0.74	0.72	19.3	14.2	0.10	0.10	11.5	11.9	.007
Model 7: +variant +methyl. score + all mRNA expr.	313	0.69	0.64	12.4	3.6	0.75	0.73	21.0	15.0	0.07	0.09	8.6	11.4	.049
Model 8: +variant +methyl. score + all methyl. sites	491	0.72	0.67	15.3	6.9	0.76	0.73	21.9	16.1	0.03	0.06	6.6	9.2	.198
	313	0.64	0.62	7.8	2.4	0.74	0.72	19.3	14.0	0.10	0.09	11.5	11.6	.005
Model 9: +variant +methyl. score + all methyl. sites +all mRNA expr.	313	0.64	0.63	7.8	2.5	0.74	0.72	19.3	14.1	0.10	0.09	11.5	11.5	.005
Average											0.09		11.9	

Note: Models including early symptoms/diagnoses were compared to otherwise equally specified models not including symptoms/diagnoses. Reported are apparent (ap) and optimism-corrected (oc) AUC/R2. An increase is calculated for performance, and an (informal) average is calculated. Model numbers refer to Figure 2, Tables S11 and S12.

[a]*p*-Value from test comparing paired ROC curves (DeLong) with and without symptoms/diagnoses included.

TABLE 4 Increase in performance due to mRNA data (qPCR PASTURE cohort).

Time	AUC without qPCR data (ap/oc)	AUC with qPCR data (ap/oc)	Increase AUC (ap/oc)	R2 without qPCR data (ap/oc)	R2 with qPCR data (ap/oc)	Increase R2 (ap/oc)	Selected genes	Estimate (delta-ct)
Cord blood	0.670/0.639	0.670/0.639	0/0	0.083/0.022	0.083/0.022	0/0	None	
Year 1	0.679/0.642	0.679/0.642	0/0	0.096/0.024	0.097/0.024	0/0	None	
Year 4	0.715/0.687	0.715/0.687	0/0	0.131/0.063	0.131/0.063	0/0	None	
Year 6	0.708/0.676	0.789/0.717	0.081/0.041	0.120/0.048	0.224/0.088	0.104/0.04	SregionIgE	−0.268
							SOCS2	−0.046
							IL10	0.069
							TNFSF13B	0.028
							LY96	0.020

Note: Optimism-corrected AUC for models without/with mRNA expression data included. Variable selection within LASSO logistic regression. All time points include variables sex, maternal asthma, variant score and farming as unpenalized covariates. Individuals with more than 15% missing were removed, followed by mean value imputation. *N* (cord blood, year 1, year 4, year 6): 702, 592, 578 and 574. A uniformly defined asthma outcome was used for all time points, as defined in methods using information up to age 11 years. mRNA expression qPCR data were specific for each time point. Delta-ct scale is indirectly proportional to expression level.

switching to a more restrictive asthma definition. We did not observe substantial differences in the variant score effect between farm and non-farm children (Section S4, Tables S17–S19, Figure S5).

4 | DISCUSSION

Clinical work-up for childhood asthma includes several risk factors and comorbidities, however, individual risk scores were suboptimal[44] and have demonstrated restricted performance for clinical practice.[11] Thus, although prognostic scores are mentioned in clinical guidelines, they are not part of specific recommendations. Nevertheless, there is consensus to integrate different multi-omics levels and clinical data to generate individual risk predictions.[45]

Using three longitudinal birth cohorts, we in parallel integrated both epidemiological and molecular factors quantifiable at birth in a prediction score for subsequent childhood asthma. Yet, molecular markers did not add substantially to predictive performance to reach clinical relevance – neither alone nor combined. Considering symptoms/diagnoses (wheeze, AD, FA) during early manifestation, a higher level of prediction performance was observed, as previously seen in other studies,[6,46,47] however, too low for clinical application.

4.1 | Components of the prediction score

We integrated sex, maternal asthma and environment as fixed components in the score. Other factors influencing risk (e.g. caesarean section, breastfeeding, antibiotics in pregnancy, older siblings) were considered, but not included due to small or highly heterogeneous effect sizes or low prevalence.[48]

We combined the strongest effects from previous GWAS on childhood asthma[14] into a variant score as objective, independently derived genetic component. Our findings of modest predictive abilities of variants to predict asthma were similar to others[19,49] and might reflect the phenomenon of missing heritability, however, contributed independently to epidemiological or molecular components present in the model.

Using a similar methodology, we integrated the results of the largest asthma-EWAS using cord blood[18] into a methylation score. For methylation data, no best practice exists for pre-processing (including QCs, outlier-detection, normalization of signal, batch effect elimination). As addressed in our sensitivity analyses, results are highly dependent on these steps hampering the use of externally derived effects. While the association of single CpG sites to childhood asthma in our cohorts was in accordance with the meta-analysis,[18] the low observed effect sizes anticipated their limited potential for prediction. This was reflected in the low improvement of performance when adding methylation data and observed before, although with different materials.[19] The methylome-wide screening identified additional CpGs, however with too limited predictive power.

Contrary to the above components, no applicable studies linking cord blood mRNA expression to childhood asthma were

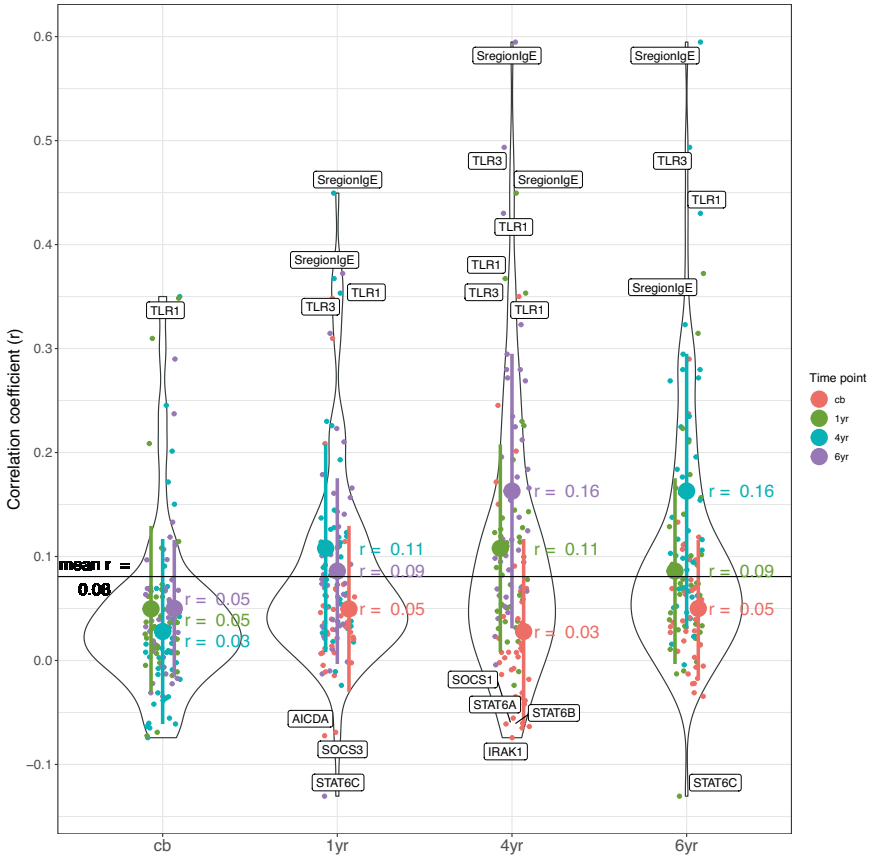


FIGURE 3 Temporal correlation of mRNA expression. Correlations were calculated on 40 genes over 4 time points (6 pairs of time points). The horizontal line shows the overall average (mean) correlation of mRNA expressions of all pairs of time points. Separate average correlation of genes between each pair of time points is indicated by coloured points and ranges (mean ± SD). Correlations above $r = .35$ and below .05 were labelled with gene names. $N = 1084$ distinct underlying PASTURE individuals involved in calculations (cb, cord blood).

available. However, association studies of asthma and mRNA expression in childhood helped to define a prioritized gene-set. Additionally, an exhaustive transcriptome-wide approach was incorporated. After pre-processing of RNASeq and NanoString-derived data, unified analysis with enhanced power was possible. Prioritizing genes with known asthma links did not improve performance over the transcriptome-wide approach. We concluded that asthma-related signals in those genes were not yet pronounced at birth, also confirmed in our qPCR data analysis showing that the association of asthma and gene expression increases with increasing age of the children (1, 4.5 and 6 years). However, single genes contributing to prediction performance were identified, but like other molecular components, absolute improvement was small and independent experiments confirming differential expression are required to allow comprehensive biological interpretations. Attempts to replicate at least some of our reported prediction models in an independent cohort were hampered by the absence of any national or international birth cohort with comparable biomaterial, identical omics assessment, and outcome data of childhood asthma at comparable age as available in our study. As a remedial measure, we therefore selected evidence-based variables with well-replicated results from the literature as far as possible. Overall, a biological link between DNA methylation and mRNA expression at birth and asthma diagnosis at school age may be challenging to detect and quantify in blood cells reflecting systemic immune regulation.

Regardless of molecular markers, the inclusion of symptoms/diagnoses yielded the most substantial increase in predictive power, aligning with most proposed scores emphasizing first symptoms (e.g. wheeze characteristics) and allergic comorbidities (rhinitis, AD)/(s)IgE testing.[12] Few studies integrated single biomarkers (exhaled breath condensate/volatile organic components)[50] and genomic/methylation markers[19,49] to improve prediction. Consistent with our results, relevant enhancements were not observed, with AUC ranging from 0.66 to 0.87 and 0.62 to 0.83, when externally validated.[51] More specifically, the only study that reached an AUC over 0.8 in external validation, was based on small numbers (28 asthmatics).[51,52] Also improvement of existing scores with molecular markers proved difficult.[12] Several factors may be responsible for the small effects of molecular markers: Prediction at birth may be too early,[53] and central underlying disease processes are not yet stably expressed. Early developmental maturation and asthma onset differ chronologically, requiring in-depth longitudinal analysis. Capturing the complex interplay and dimensionality of genetics, epigenetics and mRNA expression is statistically challenging and regulatory systems as metabolic responses through microbial influences may be important to be implemented.

4.2 | Strengths and limitations

To the best of our knowledge, this is the first study to combine genetic, methylation, and mRNA expression layers with common

asthma risk factors for predictive modelling. We used appropriate prior knowledge about risk factors for childhood asthma in contrast to re-estimate effect sizes. This helped us avoid modelling cohort-specific features and reserved power for molecular components lacking effect parameters. The model-derived predictions are thus more immediately generalizable.

Performance measures, when assessed with data already used for model development (apparent performance), are generally higher compared to assessments with new data. We addressed this by transparently reporting and focusing on optimism-corrected performance, using bootstrapping for internal validation.[54] These measures are often missing or incorrectly performed in studies, although highly recommended for prediction.[55]

One limitation of many studies including ours, despite application in three birth cohorts, is sample size. As a consequence, theoretical statistical considerations limit model complexity,[56] which might be needed to uncover biological structure. Integrating non-linear effects and interactions (e.g. gene environment) was infeasible due to the data-intensive nature of selecting from high-dimensional variables, estimating effect size, and assessing predictive performance.[57] Observing low predictive performance may reflect all combinations of low biological signals, too little power, or an inadequate modelling approach. Reporting sensitivity/specificity among others as threshold-specific measures as well as net benefit[58] would be useful as a second step after achieving sufficient score performance. Due to the observed low overall performance, we refrained from reporting these further detailed measures, crucial for assessing clinical relevance. Atopic sensitization for all cohorts and biological measurements beyond peripheral blood cells, for example, at local tissue sites (nose, lung) were not available, as invasive biomaterial sampling was impractical in these prospective cohorts. Furthermore, we restricted our analysis to a general asthma phenotype.

4.3 | Future directions

Predictive modelling would largely benefit from consent on canonical variables to include. The standard as with GWAS, to pool cohorts and report meta-analyses, requires realization also for other omics data. Ideally, the establishment of consortia that ensure uniform procedures for data collection, outcomes/variable definition, evaluation and transparent reporting should be promoted. This includes harmonizing asthma definitions[59] and improving statistical analyses/reporting.[51]

From clinical perspective, more specific asthma sub-phenotyping including allergic sensitization and conventional markers (sIgE, eosinophils, FeNO) may represent superior prediction endpoints than 'any asthma'. Subgrouping of individuals into endotypes which account for the heterogeneous nature of asthma[60] might overcome the limited predictability observed in this approach. We believe a crucial initial step involves conducting traditional association analyses with these objectively defined endotypes. This is essential in establishing the foundation for prediction models. Both, cohorts characterized in depth with omics over time and conventional markers delineating distinct pathophysiology, may offer promising tools for successful and generalizable prediction during onset of precursor symptoms. Investigation of non-invasive surrogate tissues[61] for lung/epithelial cells and purified cells may represent a compromise to overcome the highly cell-specific nature of DNA methylation and more systemic role of peripheral blood.

4.4 | Summary

Asthma prediction models, including ours, demonstrated moderate performance. Although we integrated numerous common risk factors extended by new molecular markers at birth, this score remains insufficient for clinical application. Yet, this approach is relevant and can be translated to future studies. Symptoms at early manifestation together with addition of molecular markers at an age, where more stable pattern may be established, can offer new strategies for early identification of an increased asthma risk. Further development might consider additional markers such as microbiota, metabolomics and proteomics. Prediction tools based on random forests or support vector machines offer more flexibility and make fewer structural assumptions than classical methods. Still, they cannot overcome limitations due to low sample sizes and multiplicity of analysis. Multiple omics and clinical data should be measured simultaneously, at a currently unavailable scale, followed by standardized evaluation. Overarching funding opportunities including interdisciplinary and international research consortia are key to realizing this challenging option for future asthma prediction and possibly, at some point, patient-tailored prevention.

AUTHOR CONTRIBUTIONS

AB performed data analysis and interpretation and wrote the manuscript, KU participated in study coordination and wrote the manuscript, JKE participated in study coordination and provided input to study design and data collection, MS provided input on genetic data and statistical support, KL contributed to study design and NanoString measurements, RF contributed to study design and qPCR measurements, SKu performed genotyping and DNA methylation measurements, JK provided input on NanoString measurements, MPH performed RNASeq measurements and interpretation; KK was involved in clinical interpretation. SKr, KB, EH, MK, MD, GH and BS acquired funding and contributed to project conceptualisation and administration of the CHAMP study. JR, ADC, JP and RL obtained funds, set up the PASTURE birth cohort and together with MR, ESH, CB, AMK, CR and BS were responsible for data collection and management of the study. BS was responsible for financing and study design of CHAMP as well as data analysis, interpretation and the final manuscript version. All authors contributed to and approved the final manuscript.

We sincerely thank the families for their long-term participation in the studies. We acknowledge Isolde Schleich and Tatjana Netz for

technical assistance and all study nurses for recruitment and blood withdrawal.

The graphical abstract includes elements from Servier Medical Art (https://smart.servier.com/smart_image/fetus/; licence: https://creativecommons.org/licenses/by/4.0/), openclipart.org (https://openclipart.org/detail/179428/clipboard; https://openclipart.org/detail/262568/doctor-holding-clipboard-fixed-arm-and-whiter-coat) and freesvg.org (SVG ID: 80557) both with https://creativecommons.org/publicdomain/zero/1.0/ license.

AFFILIATIONS

[1]Pediatric Allergology, Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany
[2]Member of the CHildhood Allergy and Tolerance Consortium (CHAMP), LMU Munich, Munich, Germany
[3]Comprehensive Pneumology Center - Munich (CPC-M), German Center for Lung Research (DZL), Munich, Germany
[4]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[5]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[6]Institute of Pathology, Medical Faculty, LMU Munich, Munich, Germany
[7]Institute of Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany
[8]Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany
[9]Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany
[10]Department for Pediatrics, Children's Center Bethel, University Hospital OWL, Bielefeld University, Bielefeld, Germany
[11]University Children's Hospital Regensburg (KUNO), St. Hedwig's Hospital of the Order of St. John and the University of Regensburg, Regensburg, Germany
[12]Institute of Asthma and Allergy Prevention, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
[13]Department of Pediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany
[14]Biomedical Research in Endstage and Obstructive Lung Disease Biomedical Research in Endstage and Obstructive Lung Disease (BREATH), Member of the German Center for Lung Research (DZL), Hannover, Germany
[15]Excellence Cluster Resolving Infection Susceptibility RESIST (EXC 2155), Deutsche Forschungsgemeinschaft, Hannover Medical School, Hannover, Germany
[16]Children's Hospital Schwarzach, Schwarzach, Austria
[17]Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, Finland
[18]Department of Respiratory Disease, University Hospital, Besançon, France
[19]INSERM, EFS BFC, LabEx LipSTIC, UMR1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, Univ. Bourgogne Franche-Comté, Besançon, France
[20]Pediatric Allergy Department, Children's Hospital, University Hospital of Nancy, Vandoeuvre les Nancy, France
[21]EA3450 Development, Adaptation and Handicap (devah), Pediatric Allergy Department, University of Lorraine, Nancy, France
[22]UMR/CNRS 6249 Chrono-environment, University of Franche Comté, Besançon, France
[23]Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland
[24]Department of Public Health, University of Helsinki, Helsinki, Finland
[25]Christine Kühne Center for Allergy Research and Education (CK-CARE), Davos, Switzerland
[26]Division of Respiratory Medicine and Allergology, Department of Paediatrics, Inselspital, University of Bern, Bern, Switzerland
[27]Children's Hospital of Eastern Switzerland, St. Gallen, Switzerland
[28]Children's Hospital, University of Zürich, Zürich, Switzerland

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by the Federal Ministry of Education and Research (BMBF) 01GL1742A (BS, AB, KU, JKE, KK), 01GL1742B (WG, SKr), 01GL1742C (KB), 01GL1742D (EH), 01GL1742E (MK), 01GL1742F (GH); DFG SCHA-997/8-1, SCHA-997/9-1, SCHA-997/10-1, DFG SCHA 997/11-1, Hauner Verein (BS); Finnish Institute for Health and Welfare from the Academy of Finland (grant nos. 139021, 287675, 296814, 296817 and 308254), the Juho Vainio Foundation, EVO/VTR funding, Päivikki and Sakari Sohlberg Foundation, Farmers' Social Insurance Institution (Mela), The Finnish Cultural Foundation, Yrjö Jahnsson foundation, Kuopio Area Respiratory Foundation (AMK).

CONFLICT OF INTEREST STATEMENT

AB, KU, JKE, MS, KL, SKu, JK, KK, MD, JR, MR, EHS, JP and RF have nothing to disclose. MPH reports research grants by the German Science Foundation (DFG): INST 392/165-1, HA 9694/1-1 (2022–2025), HO 6046/7-1 (2023–2025). KB reports grants from Aimmune, Danone/Nutricia, DBV, Hipp, Hycor, Infectopharm, Novartis and personal fees from Aimmune, ALK, Allergy Therapeutics, Danone/Nutricia, Hipp, Hycor, Infectopharm, Kantar Health, Primus Consulting, Mylan/Meda/Mice, Nestle, Novartis and ThermoFisher outside the submitted work and participated on Monitoring/Advisory Boards of Aimmune, Allergy Therapeutics, Danone/Nutricia, Hipp, Kantar Health, Primus Consulting, Mylan/Meda/Mice, Nestle, Novartis and leadership board roles for GPA, DGAKI, AGATE, DAAB, Healthy Start—Young Family Network. EH reports grants from BMBF, KoCON Innovationsfonds, ADCompanion Innovationsfonds and ENDEMIC Biodiversity call. He received consulting fees and participated on a Data Safety Monitoring Board or Advisory Board: Sanofi, ALK, AstraZeneca. He had leadership or fiduciary roles in other boards, societies, committees or advocacy groups: DGAKI German Allergy Society, GPP Paediatric Pneumology Society and GAN German Asthma Net. SKr reports a grant from the BMBF. MK reports grants from the European Union, the BMBF, the German Research Foundation (DFG), Infectopharm, Bavarian Ministry of Education and Research, and the Bavarian Ministry of Health; consulting fees from Bionorica, Sanofi, Novartis and Bencard; Payment or honoraria for lectures from ERS, EAACI, ATS, Novartis, Glaxo, Chiesi, Sanofi, Nutricia, Hipp, Allergopharma. He is part of a patent 'Method for testing a subject thought to have or to be predisposed to asthma', European patent application 5 EP07301135.5 and received equipment support from Roche. GH received consulting fees from Sanofi. CB reports support for attending meetings and/or travel from AstraZeneca, GSK, Sanofi, Chiesi, ALK, Novartis; Leadership for G2A (French Speaking Society of Respiratory Diseases). ADC reports a contract with the French public agency ANSES as an expert in allergy and paediatric for the Human Nutrition Committee (2018–2021), grants to University Hospital

of Besançon: from Don du Souffle for the PASTURE 16-year visit, from Fondation du Souffle for the PASTURE 16-year visit, from Novartis for a study on the protective role of farm environment on asthma and cough, to EA3450 DevAH, University of Lorraine from ARAIRLOR (Association Régionale d'Aide aux Insuffisants Respiratoires de Lorraine) for a study on the protective role of farm environment on asthma and cough. CR reports consulting fees from Aimmune, honoria for lectures from ALK, support for travel from ALK and participation on the advisory board of Aimmune. RL reports grants from the Kühne Foundation. BS reports funding support outside the present manuscript from the BMBF, DFG and EU; consulting fees from GlaxoSmithKline, Novartis, Sanofi; payment/honoraria and participation on a Data Safety Monitoring Board or Advisory Board from Sanofi.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Andreas Böck https://orcid.org/0000-0002-4511-7769
Jana Kristin Eckert https://orcid.org/0000-0003-1616-7430
Michael Salvermoser https://orcid.org/0000-0001-9412-8483
Kristina Laubhahn https://orcid.org/0000-0002-0785-9964
Sonja Kunze https://orcid.org/0000-0002-9685-2675
Marc P. Hoeppner https://orcid.org/0000-0002-0279-2270
Simone Kreimeier https://orcid.org/0000-0002-4832-7833
Kirsten Beyer https://orcid.org/0000-0003-1859-0419
Eckard Hamelmann https://orcid.org/0000-0002-2996-8248
Michael Kabesch https://orcid.org/0000-0003-0697-1871
Gesine Hansen https://orcid.org/0000-0001-5185-9454
Marjut Roponen https://orcid.org/0000-0002-4442-9090
Elisabeth Schmausser-Hechfellner https://orcid.org/0000-0002-2317-3313
Cindy Barnig https://orcid.org/0000-0001-8890-3236
Amandine Divaret-Chauveau https://orcid.org/0000-0002-2492-9864
Anne M. Karvonen https://orcid.org/0000-0003-2257-2934
Juha Pekkanen https://orcid.org/0000-0002-1083-8777
Caroline Roduit https://orcid.org/0000-0002-5988-0570
Roger Lauener https://orcid.org/0000-0002-8412-606X
Bianca Schaub https://orcid.org/0000-0003-1652-8873

REFERENCES

1. Uphoff EP, Bird PK, Antó JM, et al. Variations in the prevalence of childhood asthma and wheeze in MeDALL cohorts in Europe. *ERJ Open Res*. 2017;3:2016.
2. Doğruel D, Bingöl G, Altıntaş DU, Seydaoğlu G, Erkan A, Yılmaz M. The trend of change of allergic diseases over the years: three repeated surveys from 1994 to 2014. *Int Arch Allergy Immunol*. 2017;173:178-182.
3. García-Marcos L, Chiang C-Y, Asher MI, et al. Asthma management and control in children, adolescents, and adults in 25 countries: a Global Asthma Network Phase I cross-sectional study. *Lancet Glob Health*. 2023;11:e218-e228.
4. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med*. 1995;332:133-138.
5. Biagini Myers JM, Schauberger E, He H, et al. A Pediatric Asthma Risk Score to better predict asthma development in young children. *J Allergy Clin Immunol*. 2019;143:1803-1810.e2.
6. Castro-Rodríguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med*. 2000;162:1403-1406.
7. Kurukulaaratchy RJ, Matthews S, Holgate ST, Arshad SH. Predicting persistent disease among children who wheeze during early life. *Eur Respir J*. 2003;22:767-771.
8. Caudri D, Wijga A, A Schipper CM, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. *J Allergy Clin Immunol*. 2009;124:903-910.e7–7.
9. Pescatore AM, Dogaru CM, Duembgen L, et al. A simple asthma prediction tool for preschool children with wheeze or cough. *J Allergy Clin Immunol*. 2014;133:111-118.e1–13.
10. Colicino S, Munblit D, Minelli C, Custovic A, Cullinan P. Validation of childhood asthma predictive tools: a systematic review. *Clin Exp Allergy*. 2019;49:410-418.
11. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. 2022. Accessed March 27, 2023. www.ginasthma.org
12. Castro-Rodriguez JA, Cifuentes L, Martinez FD. Predicting asthma using clinical indexes. *Front Pediatr*. 2019;7:320.
13. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363:1211-1221.
14. Demenais F, Margaritte-Jeannin P, Barnes KC, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet*. 2018;50:42-53.
15. Ferreira MAR, Mathur R, Vonk JM, et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am J Hum Genet*. 2019;104:665-684.
16. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019;7:509-522.
17. Xu C-J, Söderhäll C, Bustamante M, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med*. 2018;6:379-388.
18. Reese SE, Xu C-J, den Dekker HT, et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J Allergy Clin Immunol*. 2019;143:2062-2074.
19. Kothalawala DM, Kadalayil L, Curtin JA, et al. Integration of genomic risk scores to improve the prediction of childhood asthma diagnosis. *J Pers Med*. 2022;12:75.
20. Forno E, Wang T, Yan Q, et al. A multiomics approach to identify genes associated with childhood asthma risk and morbidity. *Am J Respir Cell Mol Biol*. 2017;57:439-447.
21. Bigler J, Boedigheimer M, Schofield JPR, et al. A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am J Respir Crit Care Med*. 2017;195:1311-1320.
22. Jiang Y, Gruzieva O, Wang T, et al. Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents. *Eur Respir J*. 2019;53:53.
23. Ma X, Wang P, Xu G, Yu F, Ma Y. Integrative genomics analysis of various omics data and networks identify risk genes and variants vulnerable to childhood-onset asthma. *BMC Med Genet*. 2020;13:123.

24. Lemonnier N, Melén E, Jiang Y, et al. A novel whole blood gene expression signature for asthma, dermatitis, and rhinitis multimorbidity in children and adolescents. *Allergy*. 2020;75:3248-3260.
25. Schaub B, Liu J, Schleich I, Höppler S, Sattler C, von Mutius E. Impairment of T helper and T regulatory cell responses at birth. *Allergy*. 2008;63:1438-1447.
26. Schaub B, Liu J, Höppler S, et al. Maternal farm exposure modulates neonatal immune mechanisms through regulatory T cells. *J Allergy Clin Immunol*. 2009;123:774-782.e5.
27. von Mutius E, Schmid S. The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy*. 2006;61:407-413.
28. Arathimos R, Granell R, Henderson J, Relton CL, Tilling K. Sex discordance in asthma and wheeze prevalence in two longitudinal cohorts. *PLoS One*. 2017;12:e0176293.
29. Lim RH, Kobzik L, Dahl M. Risk for asthma in offspring of asthmatic mothers versus fathers: a meta-analysis. *PLoS One*. 2010;5:e10134.
30. Douwes J, Cheng S, Travier N, et al. Farm exposure in utero may protect against asthma, hay fever and eczema. *Eur Respir J*. 2008;32:603-611.
31. von Ehrenstein OS, von Mutius E, Illi S, et al. Reduced risk of hay fever and asthma among children of farmers. *Clin Exp Allergy*. 2000;30:187-193.
32. Martel M-J, Rey E, Malo J-L, et al. Determinants of the incidence of childhood asthma: a two-stage case-control study. *Am J Epidemiol*. 2009;169:195-205.
33. Genuneit J, Büchele G, Waser M, et al. The GABRIEL advanced surveys: study design, participation and evaluation of bias. *Paediatr Perinat Epidemiol*. 2011;25:436-447.
34. Illi S, Depner M, Genuneit J, et al. Protection from childhood asthma and allergy in alpine farm environments-the GABRIEL advanced studies. *J Allergy Clin Immunol*. 2012;129:1470-1477.e6.
35. Schmitz R, Atzpodien K, Schlaud M. Prevalence and risk factors of atopic diseases in German children and adolescents. *Pediatr Allergy Immunol*. 2012;23:716-723.
36. Vermeulen EM, Koplin JJ, Dharmage SC, et al. Food allergy is an important risk factor for childhood asthma, irrespective of whether it resolves. *J Allergy Clin Immunol Pract*. 2018;6:1336-1341.e3.
37. Vial Dupuy A, Amat F, Pereira B, Labbe A, Just J. A simple tool to identify infants at high risk of mild to severe childhood asthma: the persistent asthma predictive score. *J Asthma*. 2011;48:1015-1021.
38. Yang IV, Pedersen BS, Liu A, et al. DNA methylation and childhood asthma in the inner city. *J Allergy Clin Immunol*. 2015;136:69-80.
39. Wang C, Li H, Cao L, Wang G. Identification of differentially expressed genes associated with asthma in children based on the bioanalysis of the regulatory network. *Mol Med Rep*. 2018;18:2153-2163.
40. Zhang N-Z, Chen X-J, Mu Y-H, Wang H. Identification of differentially expressed genes in childhood asthma. *Medicine*. 2018;97:e10861.
41. Chen R, Piao L-Z, Liu L, Zhang X-F. DNA methylation and gene expression profiles to identify childhood atopic asthma associated genes. *BMC Pulm Med*. 2021;21:292.
42. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika*. 1991;78(3):691-692.
43. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.
44. Fabiano Filho RC, Geller RJ, Candido Santos L, et al. Performance of three asthma predictive tools in a cohort of infants hospitalized with severe bronchiolitis. *Front Allergy*. 2021;2:758719.
45. Golebski K, Kabesch M, Melén E, et al. Childhood asthma in the new omics era: challenges and perspectives. *Curr Opin Allergy Clin Immunol*. 2020;20:155-161.
46. Hafkamp-de Groen E, Lingsma HF, Caudri D, et al. Predicting asthma in preschool children with asthma-like symptoms: validating and updating the PIAMA risk score. *J Allergy Clin Immunol*. 2013;132:1303-1310.
47. Wang R, Simpson A, Custovic A, Foden P, Belgrave D, Murray CS. Individual risk assessment tool for school-age asthma prediction in UK birth cohort. *Clin Exp Allergy*. 2019;49:292-298.
48. Castro-Rodriguez JA, Forno E, Rodriguez-Martinez CE, Celedón JC. Risk and protective factors for childhood asthma: what is the evidence. *J Allergy Clin Immunol Pract*. 2016;4:1111-1122.
49. Dijk FN, Folkersma C, Gruzieva O, et al. Genetic risk scores do not improve asthma prediction in childhood. *J Allergy Clin Immunol*. 2019;144:857-860.e7.
50. Klaassen EMM, van de Kant KDG, Jöbsis Q, et al. Exhaled biomarkers and gene expression at preschool age improve asthma prediction at 6 years of age. *Am J Respir Crit Care Med*. 2015;191:201-207.
51. Kothalawala DM, Kadalayil L, Weiss VBN, et al. Prediction models for childhood asthma: a systematic review. *Pediatr Allergy Immunol*. 2020;31:616-627.
52. Grabenhenrich LB, Reich A, Fischer F, et al. The novel 10-item asthma prediction tool: external validation in the German MAS birth cohort. *PLoS One*. 2014;9:e115852.
53. Olin A, Henckel E, Chen Y, et al. Stereotypic immune system development in newborn children. *Cell*. 2018;174:1277-1292.e14.54.
54. Steyerberg EW, Bleeker SE, Moll HA, et al. internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56:441-447.
55. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol*. 2015;68:134-143.
56. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II – binary and time-to-event outcomes. *Stat Med*. 2019;38:1276-1296.
57. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14:137.
58. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26:565-574.
59. Gaillard EA, Kuehni CE, Turner S, et al. Moeller A European Respiratory Society clinical practice guidelines for the diagnosis of asthma in children aged 5-16 years. *Eur Respir J*. 2021;58:58.
60. Pijnenburg MW, Frey U, de Jongste JC, Saglani S. Childhood asthma: pathogenesis and phenotypes. *Eur Respir J*. 2022;59(6):2100731.
61. Lin X, Teh AL, Chen L, et al. Karnani N Choice of surrogate tissue influences neonatal EWAS findings. *BMC Med*. 2017;15:211.
62. Thamm R, Poethko-Müller C, Hüther A, Thamm M. Allergische Erkrankungen bei Kindern und Jugendlichen in Deutschland – Querschnittergebnisse aus KiGGS Welle 2 und Trends. *J Health Monit*. 2018;3:3-18.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Böck A, Urner K, Eckert JK, et al. An integrated molecular risk score early in life for subsequent childhood asthma risk. *Clin Exp Allergy*. 2024;54:314-328. doi:10.1111/cea.14475

APPENDIX

CHAMP study group: Franziska Sattler (Pediatric Allergology, Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany), Sabine Schnadt (German Allergy and Asthma Association [DAAB], Mönchengladbach, Germany), Julia Kahle (German Allergy and Asthma Association [DAAB], Mönchengladbach, Germany), Wolfgang Greiner (Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany), Ole Marten (Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany), Young-Ae Lee (Max Delbruck Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; Charité—Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; Experimental and Clinical Research Center, a joint cooperation of Max Delbruck Center for Molecular Medicine and Charité—Universitätsmedizin Berlin, Berlin, Germany), Josefine Dobbertin-Welsch (Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany), Christina Schorlemer (Department of Pediatrics, Children's Center Bethel, University Hospital OWL, University Bielefeld, Bielefeld), Patricia Maasjosthusmann (Department of Pediatrics, Children's Center Bethel, University Hospital OWL, University Bielefeld, Bielefeld), Parastoo Kheiroddin (University Children's Hospital Regensburg (KUNO), St. Hedwig's Hospital of the Order of St. John and the University of Regensburg, Regensburg), Daniel Walter (Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany), Stephanie DeStefano (Department of Pediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Hannover, Germany).

PASTURE study group: Martin Täubel (Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland), Pirkka V. Kirjavainen (Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland; Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland), Lucie Laurent (Department of Respiratory Disease, University Hospital of Besançon Besançon, France), Marie-Laure Dalphin (Paediatrics, University Hospital of Besançon, Besançon, France), Markus J. Ege (Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany; Member of the German Center for Lung Research, DZL, Germany), Sabina Illi (Institute for Asthma and Allergy Prevention, Helmholtz Zentrum München, Neuherberg, Germany; Member of the German Center for Lung Research, DZL, Germany), Johanna Theodorou (Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany; Member of the German Center for Lung Research, DZL, Germany), Sonali Pechlivanis (Institute for Asthma and Allergy Prevention, Helmholtz Zentrum München, Neuherberg, Germany), Francesco Foppiano (Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany), Claudia Beerweiler (Dr. von Hauner Children's Hospital, University Hospital, LMU Munich, Munich, Germany) Giulia Pagani (Institute for Asthma and Allergy Prevention, Helmholtz Zentrum München, Neuherberg, Germany).