ORIGINAL ARTICLE

Modeling Wheezing Spells Identifies Phenotypes with Different Outcomes and Genetic Associates

Sadia Haider[1], Raquel Granell[2], John Curtin[3], Sara Fontanella[1], Alex Cucco[1], Stephen Turner[4,5], Angela Simpson[3], Graham Roberts[6,7,8], Clare S. Murray[3], John W. Holloway[6,7], Graham Devereux[9], Paul Cullinan[1], Syed Hasan Arshad[7,8,10], and Adnan Custovic[1]

[1]National Heart and Lung Institute, Imperial College London, London, United Kingdom; [2]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; [3]Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom; [4]Royal Aberdeen Children's Hospital National Health Service Grampian, Aberdeen, United Kingdom; [5]Child Health, University of Aberdeen, Aberdeen, United Kingdom; [6]Human Development and Health and [10]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; [7]National Institute for Health Research Southampton Biomedical Research Centre, University Hospitals Southampton National Health Service Foundation Trust, Southampton, United Kingdom; [8]David Hide Asthma and Allergy Research Centre, Isle of Wight, United Kingdom; and [9]Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

ORCID IDs: 0000-0001-8393-5060 (S.T.); 0000-0003-2252-1248 (G.R.); 0000-0001-9998-0464 (J.W.H.); 0000-0001-5218-7071 (A. Custovic).

Abstract

Rationale: Longitudinal modeling of current wheezing identified similar phenotypes, but their characteristics often differ between studies.

Objectives: We propose that a more comprehensive description of wheeze may better describe trajectories than binary information on the presence/absence of wheezing.

Methods: We derived six multidimensional variables of wheezing spells from birth to adolescence (including duration, temporal sequencing, and the extent of persistence/recurrence). We applied partition-around-medoids clustering on these variables to derive phenotypes in five birth cohorts. We investigated within- and between-phenotype differences compared with binary latent class analysis models and ascertained associations of these phenotypes with asthma and lung function and with polymorphisms in asthma loci 17q12–21 and *CDHR3* (cadherin-related family member 3).

Measurements and Main Results: Analysis among 7,719 participants with complete data identified five spell-based wheeze phenotypes with a high degree of certainty: never (54.1%),
early-transient (ETW) (23.7%), late-onset (LOW) (6.9%), persistent (PEW) (8.3%), and a novel phenotype, intermittent wheeze (INT) (6.9%). FEV_1/FVC was lower in PEW and INT compared with ETW and LOW and declined from age 8 years to adulthood in INT. 17q12–21 and *CDHR3* polymorphisms were associated with higher odds of PEW and INT, but not ETW or LOW. Latent class analysis- and spell-based phenotypes appeared similar, but within-phenotype individual trajectories and phenotype allocation differed substantially. The spell-based approach was much more robust in dealing with missing data, and the derived clusters were more stable and internally homogeneous.

Conclusions: Modeling of spell variables identified a novel intermittent wheeze phenotype associated with lung function decline to early adulthood. Using multidimensional spell variables may better capture wheeze development and provide a more robust input for phenotype derivation.

Keywords: wheezing phenotypes; asthma; latent class; genetics; 17q12–21

At a Glance Commentary

Scientific Knowledge on the Subject: Longitudinal modeling of current wheezing identified similar phenotypes, but their characteristics often differ between studies.

What This Study Adds to the Field: Transformation of binary wheeze data into a set of multidimensional variables better captures the temporal characteristics of wheeze development and provides a more robust input for phenotype derivation. Modeling using multidimensional variables of wheezing spells identified a stable and consistent architecture of wheezing illness, including a novel intermittent phenotype associated with early lung function decline to early adulthood. Different wheezing phenotypes are underpinned by unique mechanisms and genetic associates.

Wheeze in most children remits by school age but in others may persist, with or without periods of remission. Over the past decades, a substantial effort has been devoted to understanding the heterogeneity of childhood wheezing illness, using both hypothesis-driven approaches, in which phenotypes are specified *a priori* based on clinical insights (1), and data-driven approaches, which incorporate a variety of multivariate statistical and machine learning methodologies (2). The latter have largely used latent class modeling, such as latent class analysis (LCA), in which repeated information of wheeze presence is used to uncover the temporal patterns over a specified time interval (3–15). These different symptom patterns may indicate distinct causes and biological mechanisms (16, 17), and their discovery may facilitate stratified treatment (18). However, to facilitate the identification of genetic associates and

underlying mechanisms, phenotypes should be internally homogeneous and consistent between different populations and studies.

The number of phenotypes reported in previous analyses that used LCA varied by study, but four were identified in all cohorts (19): never or infrequent wheeze (NWZ), early-transient (ETW), late-onset (LOW), and persistent wheeze (PEW). Some analyses identified one or two further "intermediate" phenotypes (3, 4, 20), which mostly arose from transient or late-onset patterns (21). However, although phenotypes in different studies are usually designated with the same name, they often differ in temporal trajectories, distributions within a population, and associated risk factors (19, 22). These differences are in part a consequence of the sample size and the timing and frequency of data collection (21). Furthermore, the confidence with which individuals are assigned to a phenotype varies across phenotypes, and a substantial number of children in such analyses are classified imprecisely (e.g., individuals with identical wheeze patterns may be assigned to different phenotypes, or individual trajectories may not follow wheeze patterns suggested by the phenotype label [13, 21, 23]).

We propose that within-class heterogeneity and inaccurate allocation of individual children may, in part, be responsible for a lack of consistent associations of discovered phenotypes with risk factors (24) and may adversely impact the ability to identify phenotype-specific genetic associates and underlying mechanisms. We hypothesize that incorporating a more comprehensive description of wheeze may better describe wheeze trajectories and derive more within-phenotype homogeneity to facilitate a better understanding of their differing etiology. To address our hypothesis, we drew on research in other fields, specifically the "spells" approach pioneered in the social sciences research on poverty dynamics (25–28), to move from the point prevalence of current wheeze to a dynamic approach that takes into account the duration of wheezing spells, their temporal sequencing, and the extent of persistence and recurrence (further details

can be found in the online supplement). To this end, we first developed a set of multidimensional variables to describe more comprehensively the temporal variation of wheeze and then applied a clustering approach based on the partition-around-medoids (PAM) algorithm (29) on these variables. We then investigated variation within and between phenotypes from binary (LCA) and indicator-based (PAM) models to ascertain whether we achieved increased within-phenotype homogeneity and investigated the associations of the derived clusters with early-life factors and asthma-related outcomes in adolescence. Finally, we tested the hypothesis that phenotypes defined using this approach have distinct genetic associates by investigating their associations with the known asthma loci (17q12–21 and *CDHR3*).

Methods

Study Design, Setting, and Participants

The Study Team for Early Life Asthma Research (STELAR) consortium (30) brings together five UK population-based birth cohorts: ALSPAC (Avon Longitudinal Study of Parents and Children) (31), Ashford (32), IOW (Isle of Wight) (33), and SEATON (Aberdeen) (34) cohorts, and the MAAS (Manchester Asthma and Allergy Study) (35). The cohorts are described in detail in the online supplement. All studies were approved by research ethics committees. Informed consent was obtained from parents, and study participants gave their assent or consent when applicable. Data were harmonized into the web-based knowledge management platform to enable joint analyses (30).

Data Sources and Definition of Variables

Validated questionnaires were completed on multiple occasions from infancy to adolescence (23). The cohort-specific time points and sample sizes are shown in Table E1 in the online supplement. For the analyses of pooled data, we defined epochs

Correspondence and requests for reprints should be addressed to Adnan Custovic, M.D., Ph.D., National Heart and Lung Institute, Imperial College London, Room 220, Norfolk Place, St. Mary's Campus Medical School, Department of Paediatrics, London W2 1PG, UK. E-mail: a.custovic@imperial.ac.uk.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

based on the data availability at shared time points across cohorts: infancy (0.5–1 yr); early childhood (2–3 yr); preschool to early school (4–5 yr); middle childhood (8–10 yr); and adolescence (14–18 yr) (23). For each child, we derived six wheeze variables:

1. Age of the first episode.
2. Age of the last recorded episode.
3. Total number of separate records over the observation period.
4. Duration of the longest spell based on the number of consecutive records of wheeze.
5. Total number of separate wheeze spells.
6. Spell type: a categorical variable defined as 0 = no wheeze, 1 = single spell, and 2 = intermittent spells (defined as at least two nonconsecutive spells of wheeze of any length).

An illustrative example of the derivation of the variables is shown in Table E2.

We performed spirometry in adolescence in all cohorts and ALSPAC and MAAS on at least three follow-ups from school-age to early adulthood. We recorded FEV_1 and FVC and expressed data as z-scores for each population.

Skin testing was performed in early to midschool-age in all cohorts and on six follow-ups in MAAS. The definitions of all variables can be found in the online supplement.

Statistical Analysis

We analyzed pooled data from participating cohorts. Figure E1 provides an overview of the analytical steps. A detailed description is provided in the online supplement.

Wheeze phenotypes from infancy to adolescence from six derived variables. To derive longitudinal wheeze patterns captured by the multidimensional variables, we used the PAM (29) algorithm coupled with the Wishart distance for mixed data (36), initially among 7,719 participants with complete data on wheezing at all five time points. To investigate whether our findings were influenced by missing data, we adopted the framework of Basagaña and colleagues (37), which integrates multiple imputations (38) into cluster analysis, and
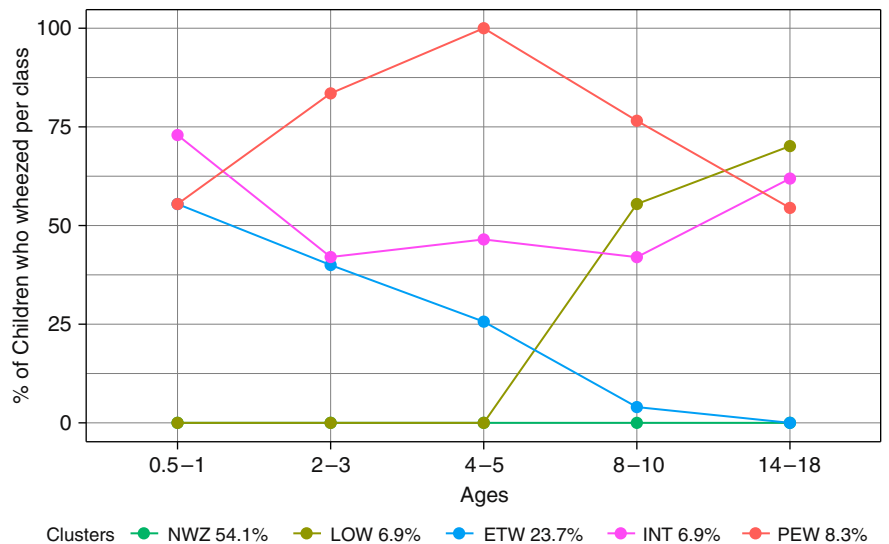

Figure 1. Trajectories of five wheeze classes were obtained with the partition-around-medoids algorithm: percentage of participants with reported wheezing in each time interval in the five cohorts. ETW = early transient; INT = intermittent; LOW = late onset; NWZ = never wheeze; PEW = persistent.

applied it to data of 15,848 participants with at least two observations.

Comparison of wheeze phenotypes derived using binary LCA and spell PAM approaches. We first repeated analyses from our previous study, which used LCA to identify five wheeze phenotypes in the same 7,719 participants (never or infrequent, preschool remitting, midchildhood remitting, persistent, and late-onset) (23), and assigned participants to phenotypes according to the maximum posterior probability. We then compared the within-class homogeneity of both models. We checked the stability of cluster allocations using the adjusted Rand index (39) and plotted the magnitude of transitions of phenotype membership between models using alluvial plots.

Association of spell-based PAM phenotypes with early-life risk factors and clinical outcomes in adolescence. We used multinomial logistic regression to ascertain early-life associates of each PAM phenotype and examine their relationship with doctor-diagnosed asthma and asthma medication use in adolescence; results are reported as relative risk ratios with 95% confidence intervals (CIs). Associations with lung function (z-scores for FEV_1, FVC, and FEV_1/FVC adjusted for height, age, and sex) were investigated using linear regression. Models were adjusted for potential confounders, including maternal history of

asthma, maternal smoking, and low birth weight.

Genetic associates of spell-based PAM phenotypes. We investigated the association of derived clusters with 17q12–21 SNPs (Table E3) and *CDHR3* SNP rs6967330 (40). We selected one representative 17q12–21 SNP per linkage disequilibrium block, leaving rs7216389, rs4795408, and rs3894194 in the final analysis. We tested the additive model using multinomial logistic regression.

Results

Characteristics of the Study Population

Of 7,719 children with complete data on wheezing, 50.4% were male. At the follow-up in adolescence, 12.9% had current asthma, and 11.4% reported using asthma medication. Demographic characteristics are shown in Table E4 and wheeze prevalence in Table E5. The prevalence of current wheeze decreased from 22.8% in infancy to 13.7% in adolescence.

Wheeze Phenotypes Obtained Using Six Derived Variables and PAM Algorithm

A five-cluster solution was selected as the optimal based on statistical fit (Figure E2).

After inspection of trajectories for each cluster (Figure 1), the clusters (phenotypes) were characterized as *1*) NWZ (54.1%); *2*) ETW (23.7%); *3*) LOW (6.9%); *4*) PEW (8.3%); and *5*) INT wheeze (6.9%). The same five-class structure was evident when we modeled each cohort separately (Table E6), and the optimal solution was stable to changes in sample size (Table E7).

Impact of missing data on cluster derivation. Detailed analysis is shown in the online supplement. The optimal solution from the model using 15,848 individuals with at least two observations was very similar to that from 7,719 participants with complete data (Table E8). Children were assigned to clusters with a high degree of certainty (Table E9). There was a very high agreement between phenotype

assignment of individual children when using complete or imputed data (adjusted Rand index = 0.94); only 195 of 7,719 (2.5%) children changed phenotype allocation (Figure E3).

Comparison of Wheezing Phenotypes Derived Using Binary LCA and Spell PAM Approaches
Figure E4 shows latent classes (phenotypes) identified by LCA. Phenotypes derived
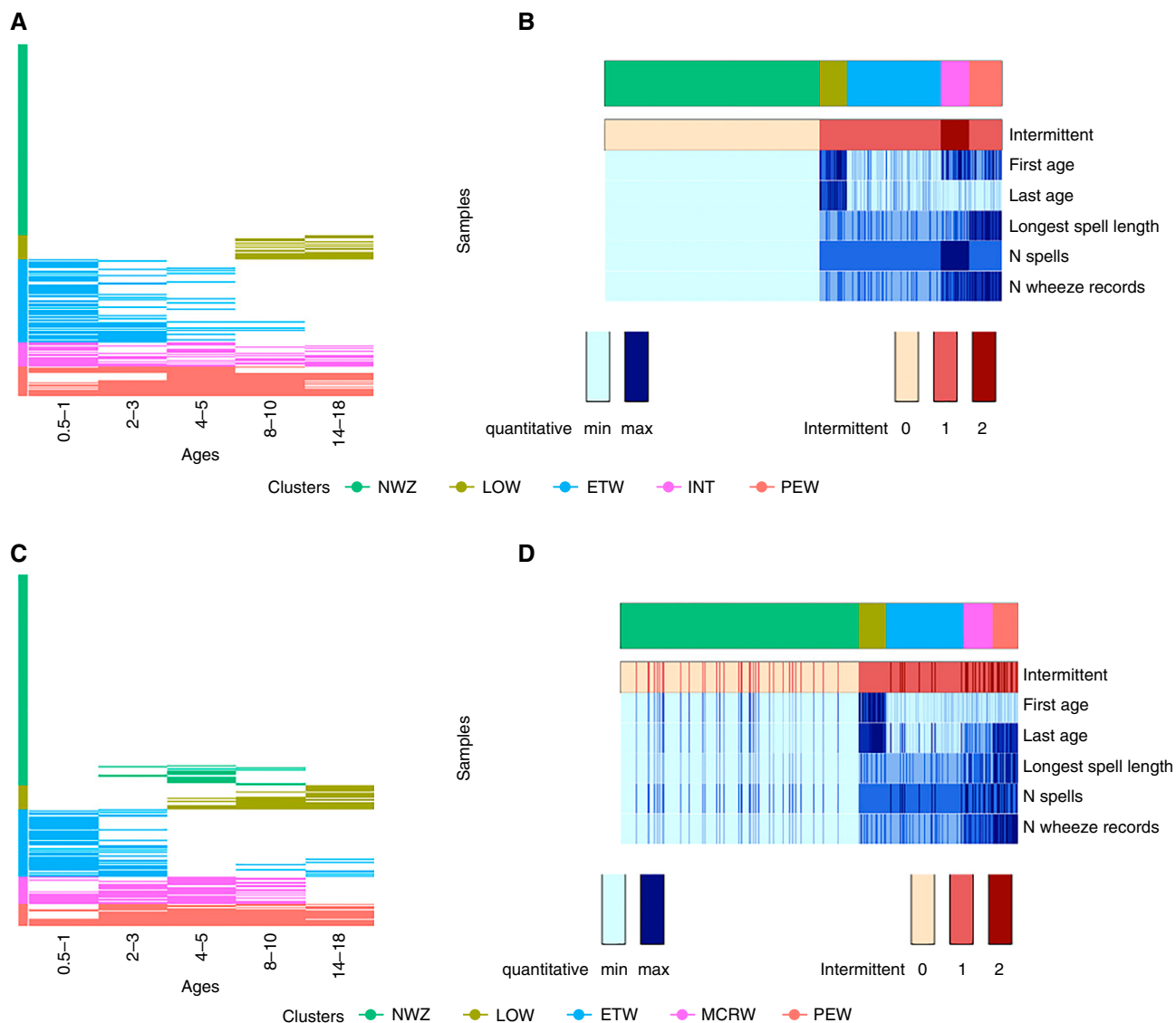
Figure 2. (*A–D*) Comparison of internal homogeneity of wheezing phenotypes derived using the spells partition-around-medoids (*A* and *B*) and binary latent class analysis approaches (*C* and *D*) among 7,719 subjects with complete data on wheezing from infancy to adolescence. (*B* and *D*) Plots are multidimensional heatmaps that show the density of the distribution of each of the six derived variables, each of which is represented as a row. The scale of the variables (quantitative and categorical) is shown at the bottom of the plot. The segments in the top bar represent each cluster and its relative size. The distribution of each indicator within each cluster is shown vertically. In *B*, for example, intermittent spells (as represented by category two for the intermittent variable) are only present in the pink INT cluster; in *D*, the latent class analysis model, intermittent wheeze is present in all classes. ETW = early transient; INT = intermittent; LOW = late onset; MCRW = midchildhood remitting wheeze; NWZ = never wheeze; PEW = persistent. For intermittent, 0 = no wheeze, 1 = single spell, and 2 = intermittent spells.

Table 1. Associations of Wheezing Phenotypes with Asthma-related Outcomes in Adolescence

	Associations with Asthma in Adolescence*				Associations with Lung Function in Adolescence*		
	Current[†] Asthma	Asthma Ever	Current[†] Asthma Medication	Asthma Medication Ever	z-Scores for FEV$_1$[‡]	z-Scores for FVC[‡]	z-Scores for FEV$_1$/FVC[‡]
Never wheeze	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Early transient	2.44 (1.84 to 3.24)	4.00 (3.45 to 4.63)	1.96 (1.48 to 2.58)	3.31 (2.92 to 3.75)	−0.103 (−0.19 to −0.02)	−0.014 (−0.10 to 0.07)	−0.151 (−0.24 to −0.07)
P value	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.021**	0.748	**<0.0001**
Intermittent	27.06 (20.44 to 35.84)	22.77 (18.23 to 28.44)	17.34 (13.17 to 22.84)	18.47 (15.21 to 22.43)	−0.168 (−0.29 to −0.05)	0.054 (−0.06 to 0.17)	−0.379 (−0.49 to −0.27)
P value	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.005**	0.37	**<0.0001**
Persistent	37.72 (29.13 to 48.85)	48.34 (38.47 to 60.74)	26.78 (20.90 to 34.32)	38.97 (32.15 to 47.24)	−0.326 (−0.45 to −0.20)	0.079 (−0.05 to 0.21)	−0.707 (−0.83 to −0.59)
P value	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.221	**<0.0001**
Late onset	35.44 (27.30 to 46.00)	17.8 (14.58 to 21.73)	16.78 (12.96 to 21.72)	22.32 (18.26 to 27.27)	−0.003 (−0.13 to 0.13)	0.159 (0.03 to 0.29)	−0.302 (−0.43 to −0.18)
P value	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.959	**0.015**	**<0.0001**

Results are from a multinomial logistic regression using children with at least two observations on wheeze (reference class: never wheeze) using weighted membership probabilities. Weights derived from probabilities of class membership across 10 imputation samples from the partition-around-medoids model. Results are reported as adjusted odds ratios with 95% confidence intervals. Bold figures indicate statistically significant differences at *P* < 0.05.
*Models adjusted for maternal history of asthma (recruitment), maternal smoking (recruitment), and low birth weight.
[†]Available at the latest follow-up (18 years in Isle of Wight, 16 years in the Manchester Asthma and Allergy Study, 15 years in Aberdeen, 15 years in Ashford, and 15 years in the Avon Longitudinal Study of Parents and Children).
[‡]Sex-, age-, and height-adjusted SD units.

using the two methods among the same 7,719 participants appeared very similar, and four appeared identical (NWZ, ETW, PEW, and LOW) (Figures 1 and E4). However, the within-phenotype structure differed substantially (Figure 2). For example, in PAM-NWZ, no participants reported wheezing at any time point (Figure 2A), whereas in LCA-NWZ, 10% reported occasional wheezing (Figure 2C). In PAM-ETW, no participants reported wheezing after age 10 years, and nobody in PAM-LOW wheezed before age 8; in contrast, in the LCA-ETW class, 8% reported wheeze up to age 18 years, and wheeze before age 10 was present among 42% in LCA-LOW.

Figures 2B and 2D and Table E10 show the distribution of wheeze variables between phenotypes from the two approaches. In PAM-LOW, the earliest observed age of wheeze onset was 7 years later than in LCA-LOW. PAM-PEW only contained children with a long single spell of wheeze, whereas subjects in the LCA-PEW also had intermittent spells.

We further investigated the differences between individual allocations to PAM and LCA phenotypes for all 32 possible wheeze

sequences across the five time points (Table E11). We did not observe any inconsistencies across cohorts in the PAM model (i.e., the same sequences were always assigned to the same cluster). In contrast, children with identical sequences were assigned by LCA to different phenotypes (e.g., "0–1–0–1-0" was assigned to three different LCA phenotypes, whereas PAM spell-based analysis always assigned this sequence to the INT phenotype).

Figure E5 shows differences in individual assignment to PAM and LCA phenotypes. One-quarter of subjects transitioned to a different phenotype. Higher stability was observed for ETW and LOW (>70%) but was relatively poor in the PEW (60%). Children in the PAM-INT cluster transitioned from all LCA phenotypes.

Finally, we applied the PAM algorithm to the binary current wheeze variable (yes or no) to investigate whether the algorithm or the transformation to spell-based variables gave rise to homogeneous phenotypes. A five-cluster solution was optimal; however, the clusters resembled LCA phenotypes (with no INT wheeze) and were structurally internally much more heterogeneous than

phenotypes obtained using the derived variables (Figure E6). Therefore, it is likely that the derived variables were, primarily, the precursor for deriving more homogeneous phenotypes.

Association of Spell-based Phenotypes with Early-Life Risk Factors and Asthma-related Outcomes

Family history, early-life factors, and environmental exposures. Univariable analyses are shown in Table E12. Table E13 shows the results of multivariable logistic regression models. Males had a higher risk of developing PEW, ETW, and INT, but not LOW. Maternal asthma and parental smoking were associated with all four clusters. Low birth weight was associated with ETW, INT, and PEW (with the strongest association with PEW) but not with LOW.

Asthma. Compared with NWZ, all four wheeze clusters were associated with a higher risk of asthma diagnosis and medication use in adolescence (Table 1). The associations were strongest for PEW and weakest for ETW (e.g., the risk of using asthma medication was approximately 14-fold

A

B
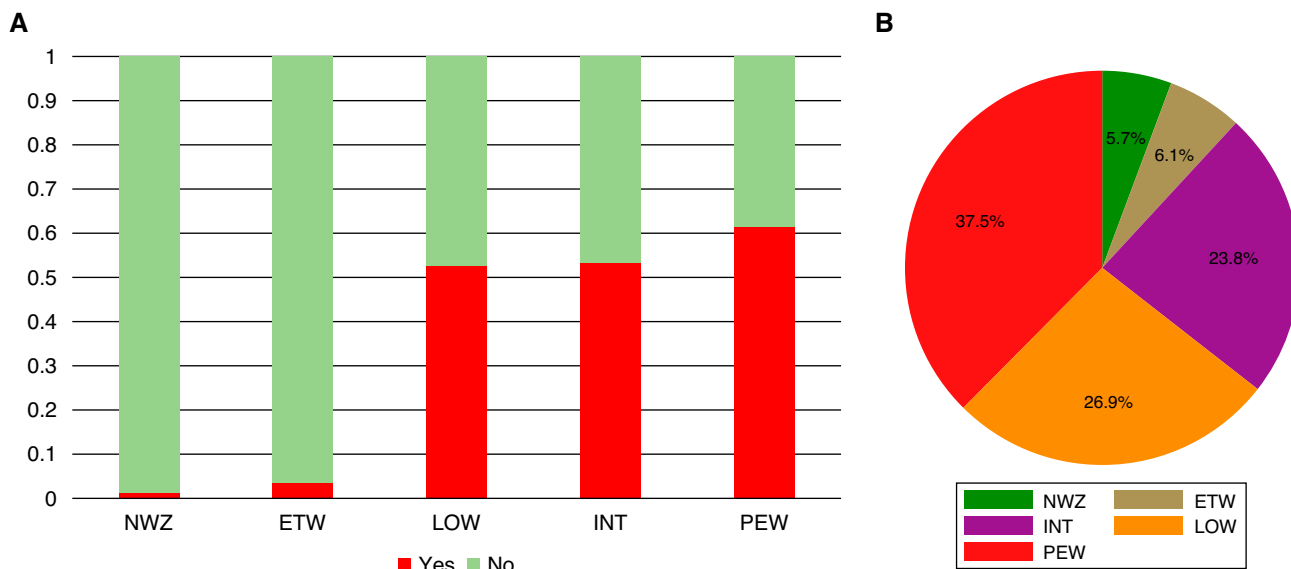
Figure 3. (A and B) The proportion of study participants with asthma diagnosis in adolescence in each partition-around-medoids wheeze phenotype (A) and the proportion of subjects with asthma diagnosis in adolescence belonging to each partition-around-medoids phenotype (B). ETW = early transient; INT = intermittent; LOW = late onset; NWZ = never wheeze; PEW = persistent.

higher for PEW than ETW). Variability in the proportion of asthmatics by spell-based phenotype and the proportion of subjects with asthma diagnosis in adolescence in each phenotype is shown in Figure 3; of note, 5.7% of children with asthma diagnosis in adolescence never reported wheezing.

Allergic sensitization. All phenotypes were associated with sensitization in early school-age (Table E13), with the magnitude of risk being higher for PEW and LOW. Trajectories of sensitization from infancy to adolescence in MAAS were almost identical in PEW, INT, and LOW and differed from those in NWZ and ETW (Figure 4) (i.e., highly concordant longitudinal sensitization patterns were associated with different wheeze phenotypes). In general, wheeze preceded sensitization in PEW and INT clusters, whereas sensitization preceded wheeze in LOW.

Lung function. FEV_1/FVC in adolescence was lower in all four wheeze phenotypes compared with children who never wheezed, with those in PEW having the lowest lung function,

markedly lower compared with NWZ (z-score, −0.71; 95% CI [−0.83 to −0.59]; $P < 0.0001$) (Table 1). FVC was similar across clusters. Longitudinal lung function was available in 6,729, 4,567, and 3,749 participants at ages 8, 15, and 24 years, respectively, in ALSPAC, and 790, 801, 630, and 504 participants at ages 8, 11, 16, and 20 years, respectively, in MAAS. FEV_1/FVC was significantly lower in all wheeze phenotypes compared with NWZ throughout the follow-up (Figure 5) and was consistently
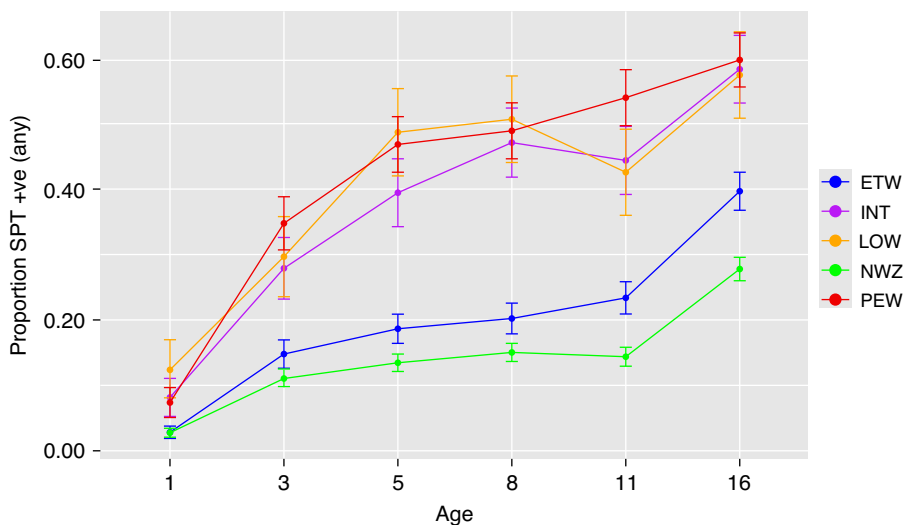
Figure 4. The proportion of children with allergic sensitization in each partition-around-medoids wheeze cluster (Manchester Asthma and Allergy Study). ETW = early transient; INT = intermittent; LOW = late onset; NWZ = never wheeze; PEW = persistent; SPT +ve = sensitized to at least one allergen based on a positive skin prick test.

Figure 5. (*A* and *B*) Lung function trajectories from early school-age to early adulthood in Manchester Asthma and Allergy Study (*A*) and Avon Longitudinal Study of Parents and Children (*B*). ETW = early transient; INT = intermittent; LOW = late onset; NWZ = never wheeze; PEW = persistent.
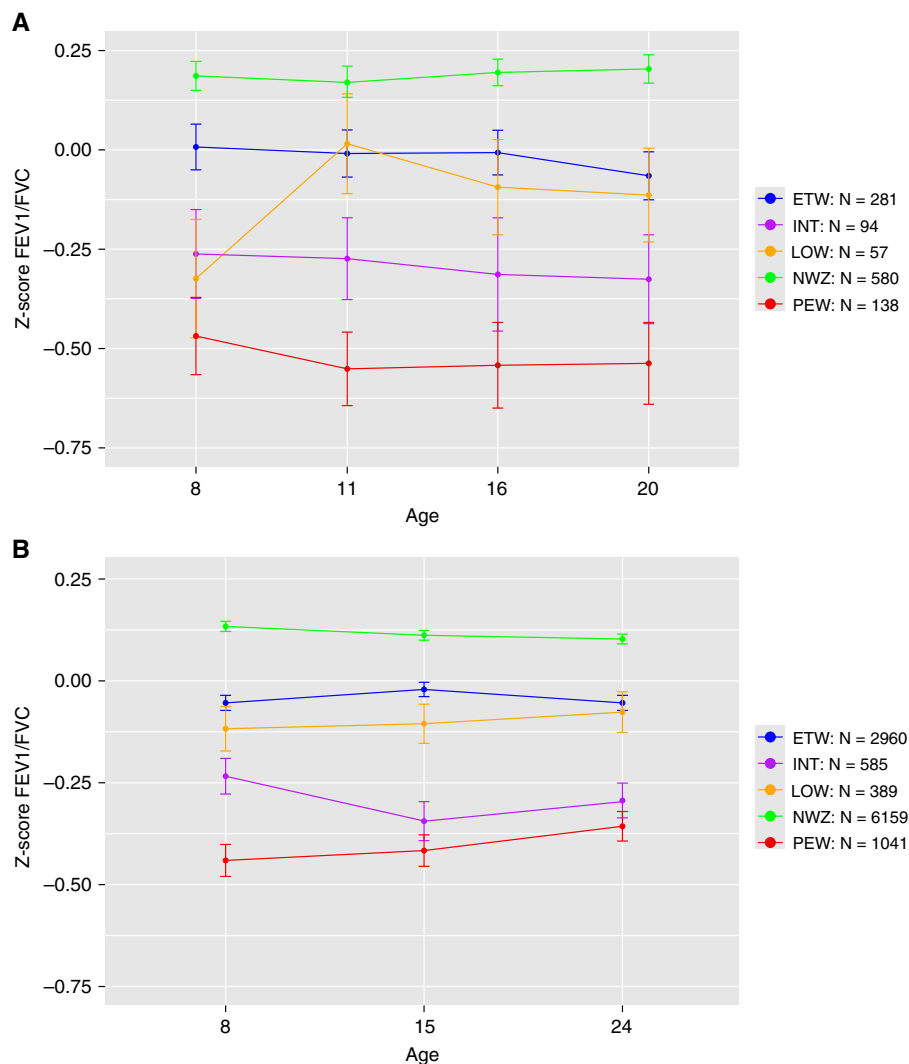
lower in PEW and INT compared with ETW and LOW (Table E14). FEV_1/FVC declined from age 8 years to early adulthood in INT, but not other phenotypes.

Association between Spell-based Phenotypes and Genetic Variants in 17q12–21 and *CDHR3*

Subjects of White European ancestry (9,655) had genotyping data and were included in the meta-analysis of genetic associations. Figure 6 shows forest plots of the associations for representative SNPs. Subgroup-level *P* values are presented in Table E15. We found strong evidence of an association

between all 17q12–21 SNPs and PEW. INT was also associated with 17q12–21 SNPs. However, we found little evidence of an association between 17q12–21 SNPs and ETW and LOW.

We found strong evidence of an association between *CDHR3* SNP rs6967330 and PEW (odds ratio, 1.45; 95% CI [1.03–2.04]) and INT (odds ratio, 1.40; 95% CI [1.04–1.89]), but there was no association with ETW and LOW clusters.

Discussion

We applied a framework that focused on wheezing spells to describe the temporal

patterns of wheeze from infancy to adolescence. Our results suggest that this approach better captures wheeze development than the presence or absence of wheezing alone and provides a more robust input for data-driven phenotype derivation. It is much more robust in dealing with missing data, and the derived clusters are stable and internally homogeneous. Our spell-based analysis applied to data from five population-based birth cohorts identified a novel wheezing phenotype, intermittent wheeze, to which ~7% of participants were assigned. FEV_1/FVC trajectory from school-age to physiological peak in early adulthood showed consistently diminished lung function in all four wheeze phenotypes determined using the spell-based approach compared with never wheezers, and in persistent and intermittent compared with transient-early and late-onset wheezing. Lung function declined from age 8 years to early adulthood in intermittent, but not other phenotypes. Finally, associations with 17q12–21 and *CDHR3* SNPs differed across wheezing phenotypes, and carriers of risk variants had significantly increased risk for persistent and intermittent, but not of transient or late-onset wheeze.

Wheezing phenotypes developed using spells appeared more clinically intuitive than those derived based on wheeze presence or absence. For example, no subjects in spell-based ETW reported wheezing after age 10 years, and nobody in LOW wheezed before age 10 years; in contrast, in the LCA-ETW, some children reported wheeze to age 18 years, and early-life wheeze was reported in some individuals assigned to LCA-LOW. In spell-based LOW, the earliest observed age of wheeze onset was 7 years later than in LCA-LOW.

Within-class heterogeneity may dilute associations with biomarkers, genetic variants, and environmental factors. Therefore, for such analyses, phenotypes derived using data-driven methods should be homogenous, and individual patterns of symptoms within each subgroup should be distinct from individuals in other subgroups. Our previous LCA showed that a substantial number of children are classified imprecisely using binary information on wheeze, particularly when an individual's posterior probability of assignment is less than 0.80 (21). Similarly, a recent US study that derived wheeze phenotypes
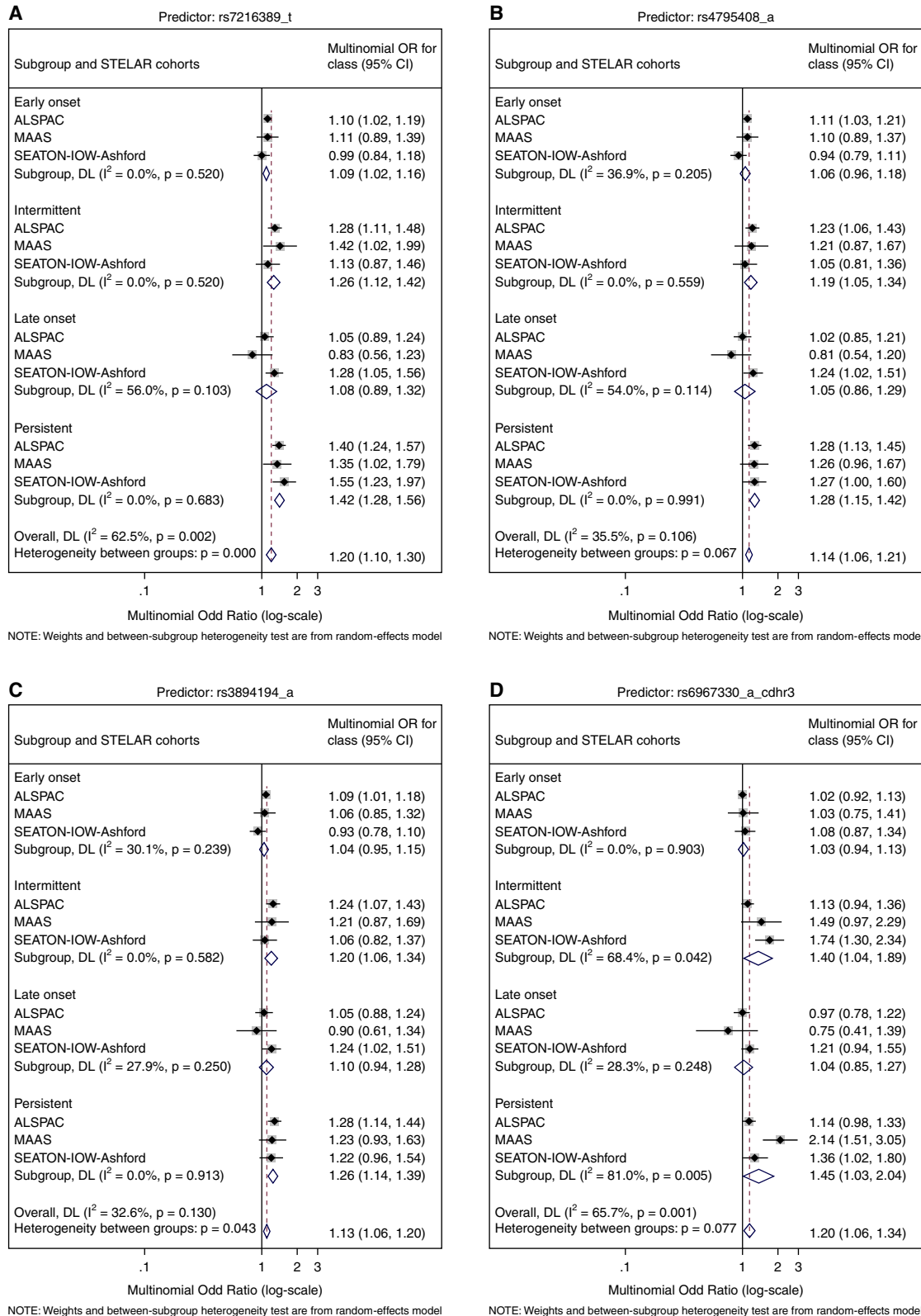
A Predictor: rs7216389_t

Subgroup and STELAR cohorts	Multinomial OR for class (95% CI)
Early onset	
ALSPAC	1.10 (1.02, 1.19)
MAAS	1.11 (0.89, 1.39)
SEATON-IOW-Ashford	0.99 (0.84, 1.18)
Subgroup, DL (I² = 0.0%, p = 0.520)	1.09 (1.02, 1.16)
Intermittent	
ALSPAC	1.28 (1.11, 1.48)
MAAS	1.42 (1.02, 1.99)
SEATON-IOW-Ashford	1.13 (0.87, 1.46)
Subgroup, DL (I² = 0.0%, p = 0.520)	1.26 (1.12, 1.42)
Late onset	
ALSPAC	1.05 (0.89, 1.24)
MAAS	0.83 (0.56, 1.23)
SEATON-IOW-Ashford	1.28 (1.05, 1.56)
Subgroup, DL (I² = 56.0%, p = 0.103)	1.08 (0.89, 1.32)
Persistent	
ALSPAC	1.40 (1.24, 1.57)
MAAS	1.35 (1.02, 1.79)
SEATON-IOW-Ashford	1.55 (1.23, 1.97)
Subgroup, DL (I² = 0.0%, p = 0.683)	1.42 (1.28, 1.56)
Overall, DL (I² = 62.5%, p = 0.002)	
Heterogeneity between groups: p = 0.000	1.20 (1.10, 1.30)

Multinomial Odd Ratio (log-scale)

NOTE: Weights and between-subgroup heterogeneity test are from random-effects model

B Predictor: rs4795408_a

Subgroup and STELAR cohorts	Multinomial OR for class (95% CI)
Early onset	
ALSPAC	1.11 (1.03, 1.21)
MAAS	1.10 (0.89, 1.37)
SEATON-IOW-Ashford	0.94 (0.79, 1.11)
Subgroup, DL (I² = 36.9%, p = 0.205)	1.06 (0.96, 1.18)
Intermittent	
ALSPAC	1.23 (1.06, 1.43)
MAAS	1.21 (0.87, 1.67)
SEATON-IOW-Ashford	1.05 (0.81, 1.36)
Subgroup, DL (I² = 0.0%, p = 0.559)	1.19 (1.05, 1.34)
Late onset	
ALSPAC	1.02 (0.85, 1.21)
MAAS	0.81 (0.54, 1.20)
SEATON-IOW-Ashford	1.24 (1.02, 1.51)
Subgroup, DL (I² = 54.0%, p = 0.114)	1.05 (0.86, 1.29)
Persistent	
ALSPAC	1.28 (1.13, 1.45)
MAAS	1.26 (0.96, 1.67)
SEATON-IOW-Ashford	1.27 (1.00, 1.60)
Subgroup, DL (I² = 0.0%, p = 0.991)	1.28 (1.15, 1.42)
Overall, DL (I² = 35.5%, p = 0.106)	
Heterogeneity between groups: p = 0.067	1.14 (1.06, 1.21)

Multinomial Odd Ratio (log-scale)

NOTE: Weights and between-subgroup heterogeneity test are from random-effects model

C Predictor: rs3894194_a

Subgroup and STELAR cohorts	Multinomial OR for class (95% CI)
Early onset	
ALSPAC	1.09 (1.01, 1.18)
MAAS	1.06 (0.85, 1.32)
SEATON-IOW-Ashford	0.93 (0.78, 1.10)
Subgroup, DL (I² = 30.1%, p = 0.239)	1.04 (0.95, 1.15)
Intermittent	
ALSPAC	1.24 (1.07, 1.43)
MAAS	1.21 (0.87, 1.69)
SEATON-IOW-Ashford	1.06 (0.82, 1.37)
Subgroup, DL (I² = 0.0%, p = 0.582)	1.20 (1.06, 1.34)
Late onset	
ALSPAC	1.05 (0.88, 1.24)
MAAS	0.90 (0.61, 1.34)
SEATON-IOW-Ashford	1.24 (1.02, 1.51)
Subgroup, DL (I² = 27.9%, p = 0.250)	1.10 (0.94, 1.28)
Persistent	
ALSPAC	1.28 (1.14, 1.44)
MAAS	1.23 (0.93, 1.63)
SEATON-IOW-Ashford	1.22 (0.96, 1.54)
Subgroup, DL (I² = 0.0%, p = 0.913)	1.26 (1.14, 1.39)
Overall, DL (I² = 32.6%, p = 0.130)	
Heterogeneity between groups: p = 0.043	1.13 (1.06, 1.20)

Multinomial Odd Ratio (log-scale)

NOTE: Weights and between-subgroup heterogeneity test are from random-effects model

D Predictor: rs6967330_a_cdhr3

Subgroup and STELAR cohorts	Multinomial OR for class (95% CI)
Early onset	
ALSPAC	1.02 (0.92, 1.13)
MAAS	1.03 (0.75, 1.41)
SEATON-IOW-Ashford	1.08 (0.87, 1.34)
Subgroup, DL (I² = 0.0%, p = 0.903)	1.03 (0.94, 1.13)
Intermittent	
ALSPAC	1.13 (0.94, 1.36)
MAAS	1.49 (0.97, 2.29)
SEATON-IOW-Ashford	1.74 (1.30, 2.34)
Subgroup, DL (I² = 68.4%, p = 0.042)	1.40 (1.04, 1.89)
Late onset	
ALSPAC	0.97 (0.78, 1.22)
MAAS	0.75 (0.41, 1.39)
SEATON-IOW-Ashford	1.21 (0.94, 1.55)
Subgroup, DL (I² = 28.3%, p = 0.248)	1.04 (0.85, 1.27)
Persistent	
ALSPAC	1.14 (0.98, 1.33)
MAAS	2.14 (1.51, 3.05)
SEATON-IOW-Ashford	1.36 (1.02, 1.80)
Subgroup, DL (I² = 81.0%, p = 0.005)	1.45 (1.03, 2.04)
Overall, DL (I² = 65.7%, p = 0.001)	
Heterogeneity between groups: p = 0.077	1.20 (1.06, 1.34)

Multinomial Odd Ratio (log-scale)

NOTE: Weights and between-subgroup heterogeneity test are from random-effects model

Figure 6. (*A–D*) Forest plots of associations of 17q12–21 SNPs (*A–C*) and *CDHR3* (cadherin-related family member 3) (*D*) with partition-around-medoids wheeze clusters. ALSPAC = Avon Longitudinal Study of Parents and Children; CI = confidence interval; DL = DerSimonian–Laird method for estimating between-study variability; IOW = Isle of Wight; MAAS = Manchester Asthma and Allergy Study; OR = odds ratio; SEATON = Aberdeen; STELAR = Study Team for Early Life Asthma Research.

using LCA found that one-third of subjects had a posterior probability of less than 0.80 (13). Our current analysis demonstrates that when using the binary representation of wheeze, some wheeze patterns are not assigned to phenotypes with high precision, and consequently, individuals with the same longitudinal wheezing patterns can be assigned to different phenotypes. The intermittent patterns contributed to substantial within-class heterogeneity when using binary data in both LCA and PAM models. Once the spell approach isolated these intermittent patterns, ETW, LOW, and PEW were more internally homogeneous, and a novel INT cluster emerged.

Our previous analysis in the same study population showed that data imputation has a major impact on the assignment of individual participants to different phenotypes in LCA (e.g., ~40% of children switched from early-onset middle-childhood remitting to PEW from the model with complete data to that with imputed data [23]). In contrast, in the current study, there was a remarkably high agreement between the assignment of individuals into clusters when using complete or imputed data, and only 2.5% of children changed phenotype. This is of key importance for longitudinal studies in which data missingness is inevitable and for genetic analyses in which a large sample size is essential.

The important question as to whether different longitudinal wheezing phenotypes are underpinned by unique pathophysiological mechanisms has been asked by Koppelman and Kersten (41) in an editorial following the recent finding from the CREW consortium, which investigated the association of 17q12–21 SNPs with LCA-derived phenotypes (13). In this study, contrary to the hypothesis of differential genetic associations of different wheeze phenotypes, associations between multiple 17q12–21 SNPs were similar for all LCA phenotypes, suggesting that all wheezing phenotypes have shared genetic origin in relation to this locus (13). In contrast, we found a clear differential association of genetic markers between phenotypes derived using spell-based variables. We found no

association of the SNPs in this locus with transient and late-onset wheezing, and our results do not support the notion that the 17q locus should be considered a "wheezing locus."

Both 17q21 locus and *CDHR3* are linked to differential susceptibility to infection by rhinoviruses (42, 43), and our data suggest that such susceptibility is common and important for early-onset nontransient phenotypes (both persistent and intermittent). However, most children who wheeze in early life stop wheezing by school-age (~2/3 in our dataset, all of whom clustered to spell-based ETW), and known genetic markers of susceptibility to rhinoviruses were not apparent in this group. This is consistent with recent data showing that even among children with severe recurrent preschool wheeze, ~50% had no evidence of either inflammation or infection in their lower airways (44). It is possible that diminished lung function in early childhood (as suggested by the seminal study from the Tucson cohort [45] and indirectly confirmed in one of our cohorts [46, 47]) is associated with poor growth in early childhood (48) or specific genetic susceptibility (49, 50) and is a principal cause of early-onset transient wheezing, whereas susceptibility to viruses may contribute to persistence and exacerbations. We cannot exclude that the immune response to other viruses (such as respiratory syncytial virus) may also be important in ETW (51). Our data also suggest that LOW (which in the current analysis started after age 10 yr) in most children may not be associated with susceptibility to viruses but is predominantly an allergic airway disease, as suggested by the analysis of the pattern of *in vitro* immune responses to viruses (52). In these individuals, allergen exposure may be the principal contributor to severity and exacerbations (53). However, it is important to emphasize that all wheeze phenotypes were associated with diminished lung function in adolescence and early adulthood, with the greatest impairment in PEW and INT. This is a precursor of chronic obstructive pulmonary disease (54–56), early all-cause mortality (57),

and early-onset cardiovascular, respiratory, and metabolic comorbidities (58).

We found that 5.7% of children with asthma diagnoses in adolescence belonged to the NWZ group (and a similar proportion to the ETW group). This emphasizes the heterogeneity of doctor-diagnosed asthma at the population level and the fact that children with other respiratory symptoms such as cough (even in the absence of wheezing) are diagnosed as being asthmatic.

One limitation of our study is that the population is not ethnically diverse. In addition, early-life pulmonary or airway function tests were not performed, which limits the inference to the potential role of premorbid lung function. We also acknowledge that our study was not able to investigate the relationship between wheeze treatment, disease severity, and patterns of wheeze spells. With respect to genetic analyses, further investigations are needed at a genomewide level to help distinguish mechanisms of early-life wheeze and subsequent asthma.

In conclusion, our data are consistent with the notion that in addition to shared pathophysiology, distinct wheezing phenotypes are underpinned by unique mechanisms and genetic associates. Modeling using multidimensional variables of wheezing spells identified a stable and consistent architecture of wheezing illness, including a novel intermittent phenotype associated with early lung function decline to early adulthood. We suggest that the transformation of binary data into a set of multidimensional variables may better capture the temporal characteristics of wheeze development and may provide a more robust input for phenotype derivation. ■

References

1. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ; The Group Health Medical Associates. Asthma and wheezing in the first six years of life. *N Engl J Med* 1995;332:133–138.
2. Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing asthma phenotypes using machine learning approaches. *Curr Allergy Asthma Rep* 2015;15:38.
3. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008;63:974–980.
4. Granell R, Henderson AJ, Sterne JA. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: a population-based birth cohort. *J Allergy Clin Immunol* 2016;138:1060–1070.e11.
5. Granell R, Sterne J, Savenije O, Kerkhof M, Smit H, Jongste JC, et al. Identification and replication of wheezing phenotypes using longitudinal latent class analysis. American Thoracic Society 2010 International Conference. New Orleans 2010. p. A6242.
6. Fitzpatrick AM, Bacharier LB, Guilbert TW, Jackson DJ, Szefler SJ, Beigelman A, et al.; NIH/NHLBI AsthmaNet. Phenotypes of recurrent wheezing in preschool children: identification by latent class analysis and utility in prediction of future exacerbation. *J Allergy Clin Immunol Pract* 2019;7:915–924.e7.
7. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, et al.; PASTURE Study Group. Clinical and epidemiologic phenotypes of childhood asthma. *Am J Respir Crit Care Med* 2014; 189:129–138.
8. Spycher BD, Silverman M, Pescatore AM, Beardsmore CS, Kuehni CE. Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts. *J Allergy Clin Immunol* 2013;132:1058–1067.
9. Belgrave DCM, Simpson A, Semic-Jusufagic A, Murray CS, Buchan I, Pickles A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *J Allergy Clin Immunol* 2013;132:575–583.e12.
10. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J* 2008;31:974–981.
11. Sordillo JE, Coull BA, Rifas-Shiman SL, Wu AC, Lutz SM, Hivert MF, et al. Characterization of longitudinal wheeze phenotypes from infancy to adolescence in Project Viva, a prebirth cohort study. *J Allergy Clin Immunol* 2020;145:716–719.e8.
12. Kotecha SJ, Watkins WJ, Lowe J, Granell R, Henderson AJ, Kotecha S. Comparison of the associations of early-life factors on wheezing phenotypes in preterm-born children and term-born children. *Am J Epidemiol* 2019;188:527–536.
13. Hallmark B, Wegienka G, Havstad S, Billheimer D, Ownby D, Mendonca EA, et al. Chromosome 17q12-21 variants are associated with multiple wheezing phenotypes in childhood. *Am J Respir Crit Care Med* 2021; 203:864–870.
14. Ödling M, Wang G, Andersson N, Hallberg J, Janson C, Bergström A, et al. Characterization of asthma trajectories from infancy to young adulthood. *J Allergy Clin Immunol Pract* 2021;9:2368–2376.e3.
15. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol* 2011;127:1505–1512.e1514.
16. Saria S, Goldenberg A. Subtyping: what it is and its role in precision medicine. *IEEE Intell Syst* 2015;30:70–75.
17. Belgrave D, Simpson A, Custovic A. Challenges in interpreting wheeze phenotypes: the clinical implications of statistical learning techniques. *Am J Respir Crit Care Med* 2014;189:121–123.
18. Saglani S, Custovic A. Childhood asthma: advances using machine learning and mechanistic studies. *Am J Respir Crit Care Med* 2019; 199:414–422.
19. Owora AH, Zhang Y. Childhood wheeze trajectory-specific risk factors: a systematic review and meta-analysis. *Pediatr Allergy Immunol* 2021; 32:34–50.
20. Granell R, Sterne JA, Savenije O, Kerkhof M, Smit HA, Jongste JD, et al. Identification and replication of wheezing phenotypes using

21. Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ; Stelar, Breathing Together investigators. Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol* 2019;143: 1783–1790 e1711.
22. Oksel C, Haider S, Fontanella S, Frainay C, Custovic A. Classification of pediatric asthma: from phenotype discovery to clinical practice. *Front Pediatr* 2018;6:258.
23. Oksel C, Granell R, Haider S, Fontanella S, Simpson A, Turner S, et al.; STELAR investigators, breathing Together investigators. Distinguishing wheezing phenotypes from infancy to adolescence: a pooled analysis of five birth cohorts. *Ann Am Thorac Soc* 2019;16:868–876.
24. Belgrave DC, Custovic A, Simpson A. Characterizing wheeze phenotypes to identify endotypes of childhood asthma, and the implications for future management. *Expert Rev Clin Immunol* 2013; 9:921–936.
25. Bane MJ, Ellwood DT. Slipping into and out of poverty: the dynamics of spells. *J Hum Resour* 1986;21:2–23.
26. Layte R, Whelan CT. Moving in and out of poverty: the impact of welfare regimes on poverty dynamics in the EU. *Eur Soc* 2003;5:167–191.
27. Mendola D, Busetta A. the importance of consecutive spells of poverty: a path-dependent index of longitudinal poverty. *Rev Income Wealth* 2012;58:355–374.
28. Stevens AH. Climbing out of poverty, falling back in: measuring the persistence of poverty over multiple spells. *J Hum Resour* 1999;34: 557–588.
29. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: Wiley; 1990.
30. Custovic A, Ainsworth J, Arshad H, Bishop C, Buchan I, Cullinan P, et al. The Study Team for Early Life Asthma Research (STELAR) consortium 'asthma e-lab': team science bringing data, methods and investigators together. *Thorax* 2015;70:799–801.
31. Golding J, Pembrey M, Jones R; ALSPAC Study Team. ALSPAC–the Avon longitudinal study of parents and children. I. Study methodology. *Paediatr Perinat Epidemiol* 2001;15:74–87.
32. Cullinan P, MacNeill SJ, Harris JM, Moffat S, White C, Mills P, et al. Early allergen exposure, skin prick responses, and atopic wheeze at age 5 in English children: a cohort study. *Thorax* 2004; 59:855–861.
33. Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, et al. Cohort profile: the Isle of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol* 2018;47:1043–1044i.
34. Martindale S, McNeill G, Devereux G, Campbell D, Russell G, Seaton A. Antioxidant intake in pregnancy in relation to wheeze and eczema in the first two years of life. *Am J Respir Crit Care Med* 2005;171:121–128.
35. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A, ; NAC Manchester Asthma and Allergy Study Group. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatr Allergy Immunol* 2002;13:32–37.
36. Wishart D. k-Means clustering with outlier detection, mixed variables and missing values. In: Schwaiger M, Opitz O, editors. Exploratory data analysis in empirical research. Berlin, Germany: Springer; 2003. pp. 216–226.
37. Basagaña X, Barrera-Gómez J, Benet M, Antó JM, Garcia-Aymerich J. A framework for multiple imputation in cluster analysis. *Am J Epidemiol* 2013;177:718–725.
38. Little RJA, Rubin DB. Statistical analysis with missing data, 2nd ed. New York: John Wiley & Sons Inc; 2002.
39. Hubert L, Arabie P. Comparing partitions. *J Classif* 1985;2:193–218.
40. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014;46:51–55.
41. Koppelman GH, Kersten ETG. Understanding how asthma starts: longitudinal patterns of wheeze and the chromosome 17q locus. *Am J Respir Crit Care Med* 2021;203:793–795.
42. Zhang Y, Willis-Owen SAG, Spiegel S, Lloyd CM, Moffatt MF, Cookson WOCM. The ORMDL3 asthma gene regulates ICAM1 and has multiple effects on cellular inflammation. *Am J Respir Crit Care Med* 2019;199: 478–488.

43. Basnet S, Bochkov YA, Brockman-Schneider RA, Kuipers I, Aesif SW, Jackson DJ, *et al. CDHR3* asthma-risk genotype affects susceptibility of airway epithelium to rhinovirus C infections. *Am J Respir Cell Mol Biol* 2019;61:450–458.

44. Robinson PFM, Fontanella S, Ananth S, Martin Alonso A, Cook J, Kaya-de Vries D, *et al.* Recurrent severe preschool wheeze: from prespecified diagnostic labels to underlying endotypes. *Am J Respir Crit Care Med* 2021;204:523–535.

45. Morgan WJ, Stern DA, Sherrill DL, Guerra S, Holberg CJ, Guilbert TW, *et al.* Outcome of asthma and wheezing in the first 6 years of life: follow-up through adolescence. *Am J Respir Crit Care Med* 2005;172:1253–1258.

46. Lowe LA, Simpson A, Woodcock A, Morris J, Murray CS, Custovic A; NAC Manchester Asthma and Allergy Study Group. Wheeze phenotypes and lung function in preschool children. *Am J Respir Crit Care Med* 2005;171:231–237.

47. Belgrave DC, Buchan I, Bishop C, Lowe L, Simpson A, Custovic A. Trajectories of lung function during childhood. *Am J Respir Crit Care Med* 2014;189:1101–1109.

48. Voraphani N, Stern DA, Zhai J, Wright AL, Halonen M, Sherrill DL, *et al.* The role of growth and nutrition in the early origins of spirometric restriction in adult life: a longitudinal, multicohort, population-based study. *Lancet Respir Med* 2022;10:59–71.

49. Simpson A, Custovic A, Tepper R, Graves P, Stern DA, Jones M, *et al.* Genetic variation in vascular endothelial growth factor-a and lung function. *Am J Respir Crit Care Med* 2012;185:1197–1204.

50. Simpson A, Maniatis N, Jury F, Cakebread JA, Lowe LA, Holgate ST, *et al.* Polymorphisms in a disintegrin and metalloprotease 33 (ADAM33) predict impaired early-life lung function. *Am J Respir Crit Care Med* 2005;172:55–60.

51. Raita Y, Pérez-Losada M, Freishtat RJ, Harmon B, Mansbach JM, Piedra PA, *et al.* Integrated omics endotyping of infants with respiratory syncytial virus bronchiolitis and risk of childhood asthma. *Nat Commun* 2021;12:3601.

52. Custovic A, Belgrave D, Lin L, Bakhsoliani E, Telcian AG, Solari R, *et al.* Cytokine responses to rhinovirus and development of asthma, allergic sensitization, and respiratory infections during childhood. *Am J Respir Crit Care Med* 2018;197:1265–1274.

53. Custovic A, Custovic D, Kljaić Bukvić B, Fontanella S, Haider S. Atopic phenotypes and their implication in the atopic march. *Expert Rev Clin Immunol* 2020;16:873–881.

54. Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, *et al.* Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med* 2018;6:535–544.

55. Belgrave DCM, Granell R, Turner SW, Curtin JA, Buchan IE, Le Souëf PN, *et al.* Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir Med* 2018;6:526–534.

56. Berry CE, Billheimer D, Jenkins IC, Lu ZJ, Stern DA, Gerald LB, *et al.* A distinct low lung function trajectory from childhood to the fourth decade of life. *Am J Respir Crit Care Med* 2016;194:607–612.

57. Vasquez MM, Zhou M, Hu C, Martinez FD, Guerra S. Low lung function in young adult life is associated with early mortality. *Am J Respir Crit Care Med* 2017;195:1399–1401.

58. Agustí A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med* 2017;5:935–945.