Systems biology and big data in asthma and allergy: recent discoveries and emerging challenges

Howard H.F. Tang [1,2,3], Peter D. Sly[4,5], Patrick G. Holt [4,5], Kathryn E. Holt[6,7] and Michael Inouye[1,2,3,8]

Affiliations: [1]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Australia. [2]Cambridge Baker Systems Genomics Initiative, Dept of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [3]School of BioSciences, The University of Melbourne, Parkville, Australia. [4]Queensland Children's Medical Research Institute, The University of Queensland, Brisbane, Australia. [5]Telethon Kids Institute, University of Western Australia, Perth, Australia. [6]Dept of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Australia. [7]London School of Hygiene and Tropical Medicine, London, UK. [8]The Alan Turing Institute, London, UK.

Correspondence: Howard H.F. Tang, Baker Heart and Diabetes Institute, Systems Genomics, 75 Commercial Rd, PO Box 6492, Melbourne VIC 3004, Australia E-mail: Howard.Tang@baker.edu.au

ABSTRACT   Asthma is a common condition caused by immune and respiratory dysfunction, and it is often linked to allergy. A systems perspective may prove helpful in unravelling the complexity of asthma and allergy. Our aim is to give an overview of systems biology approaches used in allergy and asthma research. Specifically, we describe recent "omic"-level findings, and examine how these findings have been systematically integrated to generate further insight.

Current research suggests that allergy is driven by genetic and epigenetic factors, in concert with environmental factors such as microbiome and diet, leading to early-life disturbance in immunological development and disruption of balance within key immuno-inflammatory pathways. Variation in inherited susceptibility and exposures causes heterogeneity in manifestations of asthma and other allergic diseases. Machine learning approaches are being used to explore this heterogeneity, and to probe the pathophysiological patterns or "endotypes" that correlate with subphenotypes of asthma and allergy. Mathematical models are being built based on genomic, transcriptomic and proteomic data to predict or discriminate disease phenotypes, and to describe the biomolecular networks behind asthma.

The use of systems biology in allergy and asthma research is rapidly growing, and has so far yielded fruitful results. However, the scale and multidisciplinary nature of this research means that it is accompanied by new challenges. Ultimately, it is hoped that systems medicine, with its integration of omics data into clinical practice, can pave the way to more precise, personalised and effective management of asthma.

Introduction

Asthma is a common yet complex disease that involves immune and respiratory dysfunction, and it is often associated with allergy. The ongoing prevalence of asthma and allergy has been linked to changes in environment and lifestyle [1, 2]. But while we know of several genetic and environmental determinants of asthma, the potential interactions between these determinants remain unclear. Furthermore, asthma and allergy are umbrella terms that describe a spectrum of disease, with unexplained heterogeneity in clinical manifestations. Finally, with the development of high-throughput technologies, we may be able to unravel this complexity, but it remains challenging to process, analyse and interpret the large volumes of biological data that emerge from these technologies. All these challenges have prompted researchers to search for new methods of inquiry more suited to these research problems.

Systems biology is a relatively recent development that addresses the growing complexity of biomedical research questions. The term was coined in the 1960s to describe mathematical modelling of physiological systems [3]. Today it embodies expertise across multiple fields, including biology, mathematics, statistics, informatics and computer science. The "systems" community is diverse and as such there is no singular definition of the term "systems biology" [4]. However, it is commonly presented as the study of biomedical problems involving complex systems and their interactions, by surveying and integrating high-volume data that may cover wide spatiotemporal scales [3]. These "big datasets" typically originate from "omics", fields of study involving high-throughput measurement of biomolecules: for instance, genomics for DNA, transcriptomics for RNA transcripts and proteomics for translated proteins (figure 1). Mathematical and computational expertise is then required to explore this high-volume data, using techniques such as dimension reduction, data- and text-mining, modified statistical analyses that account for spatiotemporal complexity and multiple testing burden, machine learning, and mathematical modelling. Further perturbation experiments may be performed, where a biological system can be disrupted (*e.g. via* receptor antagonists or gene knockouts) to identify functionally relevant elements of the system [5, 6]. Therefore, systems biology is by its very nature multi- and inter-disciplinary.

The practice of systems biology follows two approaches: an unbiased, hypothesis-free data-driven approach, where few *a priori* assumptions are made and models are learnt from the data; and a

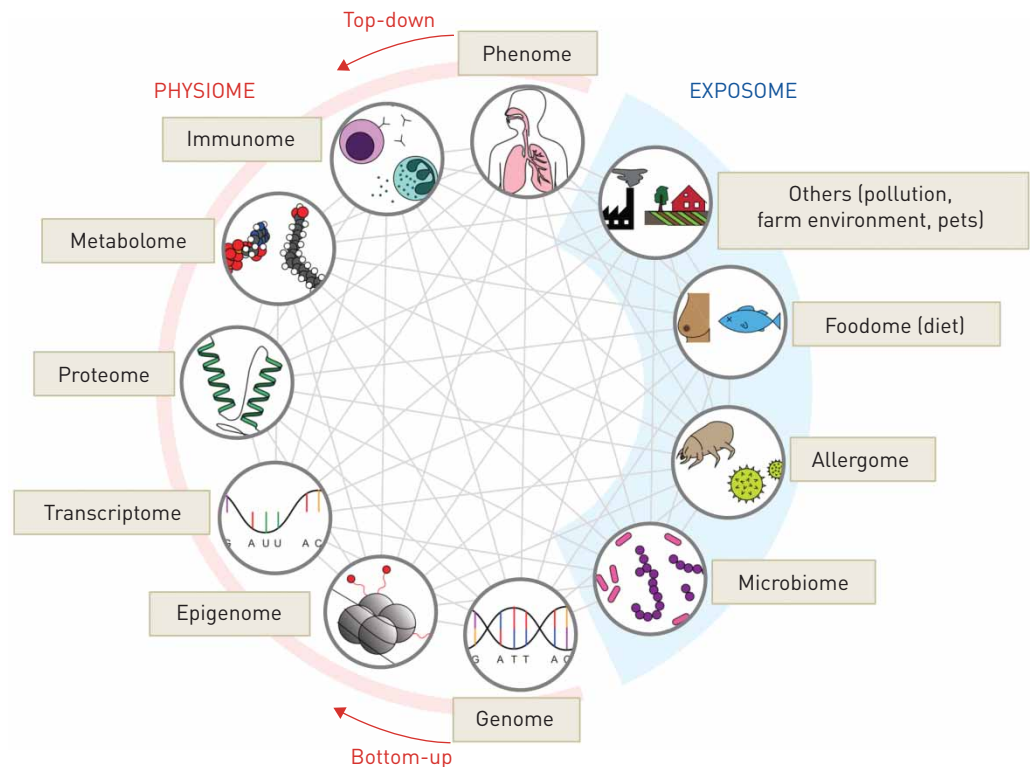
FIGURE 1 "Omics" in allergy, and their interrelationships. A depiction of the various omics that can be found in allergy and asthma research. Lines connecting the omics represent various biological relationships, associations or interactions that may exist. In systems biology, bottom-up approaches progress from the molecular scale to the macroscopic scale; *vice versa* for top-down approaches.

hypothesis-driven approach, where model design and analysis are guided by previous experiments and expert knowledge [7]. The data-driven approach is becoming increasingly popular as it can uncover new knowledge on emergent behaviour from complex data, and can be used to generate new hypotheses. In turn, hypothesis-driven studies can also be used to test those very hypotheses. For instance, data-driven approaches allow us to determine new subphenotypes of asthma, while hypothesis-driven studies will allow us to test the relationship of these subphenotypes to existing paradigms of disease (T2-driven *versus* non-T2). Furthermore, systems biology can be dichotomised into top-down *versus* bottom-up approaches, to describe the direction of enquiry in decreasing (big to small, long to short, system to components) or increasing spatiotemporal scales (small to big, *etc.*), respectively (figure 1) [8, 9].

On the surface, systems biology seems antithetical to the reductionist paradigm of old. However, systems-based approaches can produce new insights on how to proceed with reductionist experiments, and *vice versa*. In addition, there are strengths and weaknesses attached to each; while reductionist methods can oversimplify problems, their tests are more appropriate in contexts such as causal inference. Nonetheless, systems approaches are becoming indispensable to biomedical research; they allow us to better understand disease phenomena, and form the basis for precision medicine, helping us improve the screening and management of disease.

Asthma and allergy, as biomedical problems, are well suited to systems approaches. These diseases have complex pathogenesis, with multiple tiers of biological complexity, polygenicity and gene–environment interactions. Systems approaches used in asthma and allergy research include 1) discovery of disease associations within each omic field; 2) identification of relationships within and across omic fields; 3) examination of heterogeneity of disease states and phenotypes, typically by exploring the multidimensional structure of omic data *via* clustering or classification; 4) investigation of interconnections between system components in omic data by network analysis; and 5) mathematical modelling to model physiological systems or disease states, and to generate and test predictions (figure 2). Although the final approach is closest to the original formulation of systems biology, our review takes a high-level look at all approaches, with a focus on the first three.

Overview of omic findings in allergy and asthma

We begin our examination from the bottom up: from the molecular level of genomes and transcriptomes to the macroscopic level of observable phenotypes. We offer high-level summaries of recent findings at each level of profiling. As allergy-related mechanisms comprise a significant portion of asthma pathogenesis, much of the discussion involves findings related to allergic diseases at large. However, there is some exploration of the omics of nonallergic asthma.

Genomics

Asthma and allergy are highly heritable, with estimated heritability ranging from 35% to 95% [10]. In the past half-century, the quantitation of genetic variation has progressed from rough "ballpark" measurements, such as restriction fragment length polymorphisms (RFLP), to precise single-nucleotide variants or polymorphisms (SNPs) interrogated *en masse* using DNA microarrays. More recently, there has been a move towards whole-exome and whole-genome sequencing. The complexity of genetic data analysis has grown in parallel, from candidate gene studies, to genome-wide linkage studies within pedigrees, to genome-wide association analyses (GWAS) [11]. The GWAS approach is based on one-by-one association testing of thousands to millions of genetic variants across the genome with a phenotype of interest (*e.g.* asthma *versus* non-asthma status), with subsequent statistical adjustments for the multiple testing burden. The logic of adjusting for multiple testing is fundamental to many other omic-wide analyses, and is a key reason for the need for large sample sizes when dealing with such analysis.

However, while some older associations from linkage and candidate gene studies have been replicated in GWAS (*e.g.* *IL13*/*IL4* and *IL4R*), most have not. There is low concordance of significant results between these older studies and GWAS. These suggest that 1) older findings may be plagued by false positives; 2) each approach may have its own use: positional candidates from linkage studies may flag variants determining intrafamily disease risk, while GWAS flag variants determining population-wide risk; and 3) rarer or weaker gene associations may need larger sample sizes in GWAS to achieve stringent multiple testing thresholds. The latter is exacerbated by the difficulty of precise phenotyping in large population samples. However, the prevailing view in the human genetics community is that linkage and candidate gene studies of old were hampered by limitations in methodology and assumptions and frequently did not replicate in independent samples, thus many have been discounted in favour of those provided by GWAS [12].

In the past 10 years, GWAS have identified loci shared across multiple allergic phenotypes, including asthma, allergic rhinitis/hayfever, atopic dermatitis/eczema and food allergy (supplementary table S1). These probably represent genetic contributors to general allergy, and include: the human leukocyte antigen
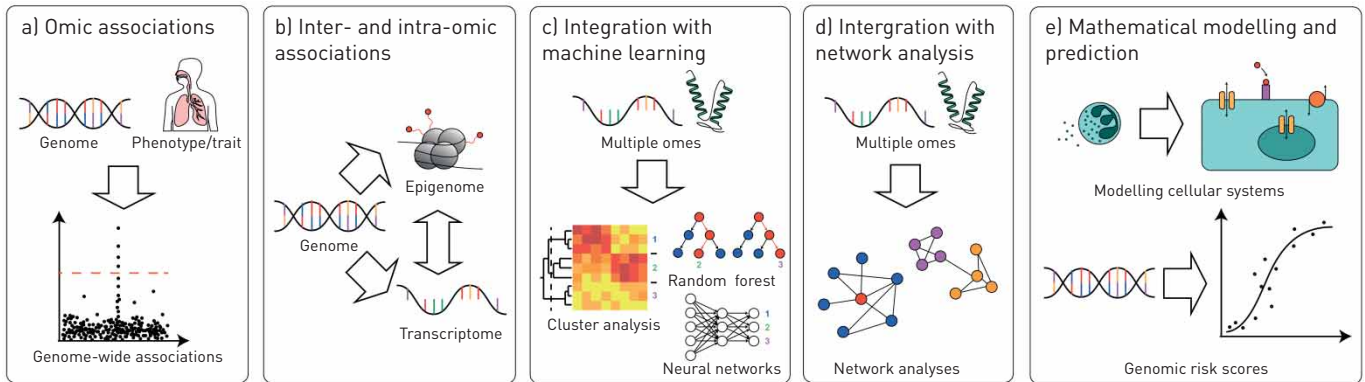
FIGURE 2 Overview of systems-based approaches to tackling research questions in allergy and asthma. The various ways in which systems biology of allergy can be interrogated: a) discovery of disease associations within each omic field of enquiry; b) identification of relationships within and across omics; c) examination of the heterogeneity of disease states or phenotypes, typically by exploring the structure of omic data *via* clustering or classification; d) investigation of interconnections between system components in omic data by network analysis; and e) mathematical modelling to model physiological systems or disease states, and to generate and test predictions. Diagrams are for illustrative purposes only and do not convey real data. a) shows a simplified Manhattan plot, with the vertical axis representing negative log p-value (statistical significance), horizontal axis representing chromosome and position and each point representing a single nucleotide polymorphism; the red line is the adjusted p-value threshold, and significant loci are marked by "peaks" extending beyond the red line.

(HLA) locus, specifically *HLA-DQ/DRB1*, *HLA-DQA1/2* and *HLA-B/C* (6p21.32-33); *C11orf30/LRRC32* (11q13.5); *IL13/RAD50* (5q31.1); *IL1RL1/IL18R1* (2q12.1); and *TSLP/WDR36* (5q22.1) [11, 13–15]. Some of these have plausible biological underpinnings linked to immune function; biomolecules such as *IL13*, *IL4*, *IL33* and *TSLP* are related to the "T2" immune response of type 2 helper T-cells (Th2) and innate lymphoid cells (ILC)2, and these are classically implicated in allergy. The HLA region encodes major histocompatibility complex class II molecules responsible for antigen presentation. Other associated genes remain uncertain in terms of pathophysiology (*e.g. WDR36*, *CLEC16A*), and require further investigation.

Due to frequent comorbidity of asthma with other allergic diseases, there is difficulty in discerning asthma-specific loci. Many loci previously thought to be unique to asthma [15] have now been found across multiple allergic diseases [16]. However, there are some loci that do appear to act specifically for certain asthma phenotypes. In particular, *ORMDL3/GSDMB/LRRC3C* (17q21.1) is linked to childhood-onset asthma [17–19]. Some loci (*e.g. TLR1/TLR6*, *ADAD1/IL2*) may be linked to Th17-related mechanisms of disease (supplementary table S1). Several studies and reviews have explored loci for asthma subphenotypes (*e.g.* aspirin-mediated and occupational asthma) and in relation to other respiratory traits (*e.g.* lung function, chronic obstructive pulmonary disease (COPD) and viral respiratory infections) [20–24]; however, these results have been inconsistent. Lately, there has been a focus on loci with ethnicity-specific effects: most of the aforementioned loci were identified primarily in European cohorts, and newer studies have begun exploring non-European populations. For instance, *PYHIN1* is significantly associated with asthma, but only in individuals of African ancestry [18]. In addition, there is an increasing focus on using admixture to map risk loci [10].

Subsequent to these findings, significant loci and their molecular products have been targeted *via* numerous pharmacological approaches. Prior to the GWAS era, numerous anti-IgE (omalizumab), anti-interleukin (IL)5 (mepolizumab), anti-IL13 (lebrikizumab), anti-IL4R (dupilmab) and anti-IL2RA (daclizumab) antibodies [25–28] had been trialled with varying degrees of success. Following GWAS discovery of new loci, other therapies (anti-IL33, IL6R, thymic stromal lymphopoietin (TSLP), *etc.*) have been tested, again with modest results [28] (supplementary table S1). Such "biologics" are currently reserved for asthma resistant to conventional forms of treatment [28, 29]. It is probable that certain biologics (*e.g.* anti-T2 cytokine therapies) are only effective in individuals whose asthma is driven by specific mechanisms (*e.g.* T2 and not T1 or Th17/ILC3). In addition, it is possible that varying efficacy may depend on patient genetics and asthma subphenotype. Indeed, GWAS for responsiveness to asthma therapy with β$_2$-agonist bronchodilators, leukotriene modifiers and steroids [30, 31] have identified loci distinct from disease-susceptibility loci. However, there have so far been few pharmacogenomic studies for the reverse: directly exploring the effects of GWAS-derived disease risk variants on treatment efficacy. Furthermore, many existing studies have been limited by lack of replication or inconsistent results. It remains a challenge to apply pharmacogenetic findings in practice, and they have yet to make a significant impact on current treatment and management.

Lately, there has been a push for risk scores based on genome-wide summary statistics. Despite the large number of novel associations discovered using GWAS, these collectively explain only a small proportion of the total heritability of asthma and allergy. The use of significant SNPs as a predictive tool for disease is often limited [32]. The existing criteria for genome-wide significance may not be sensitive for so-called "mid-hanging fruit" [33]: loci that are not genome-wide significant but still have an incremental effect on the phenotype. Recently, alternative strategies such as genomic or polygenic risk scores (PRS) have been employed to account for this missing signal. These use summary statistics from existing large-scale GWAS to generate additive scores from either a genome-wide assortment of SNPs, or from a selection of highly predictive SNPs. These have shown promise as predictive or risk-stratifying tools for other chronic polygenic diseases such as cardiovascular disease [32], but in asthma and allergy research, existing PRS have so far been limited to small subsets of genome-wide significant SNPs. BELSKY et al. [34] derived a PRS based on 17 SNPs from an asthma GWAS [17], pruned by significance and linkage disequilibrium R^2 threshold; this score was predictive for earlier asthma onset, allergy, reduced lung function and risk of childhood asthma becoming persistent into adulthood. ARABKHAZAELI et al. [35] developed a similar score for childhood allergy using 10 SNPs from a GWAS for adult allergy [36]. These methods, while interesting, may be less predictive than models that use a broader genome-wide selection of thousands to millions of SNPs, accounting for the known polygenic architectures of the diseases. For example, LEHTO et al. [37] used a genome-wide PRS, but for affective traits, to identify possible shared genetic influences between asthma and depression. It remains to be seen whether such findings replicate across multiple studies, and whether PRS can be used to reliably capture disease pathophysiology.

Transcriptomics

The transcriptome represents the entire repertoire of genes expressed in an organism or cell. Mirroring the developments in genomics, there has been a move from investigation of single-gene transcripts *via* traditional methods (*e.g.* Northern blotting), to genome-wide methods involving oligonucleotide microarrays, and most recently to RNA sequencing (RNA-seq, which involves reverse transcription to cDNA followed by deep sequencing) [38]. Transcriptomes may be determined by aligning RNA-seq reads to transcripts annotated in a reference genome, or assembling transcripts *de novo*, followed by quantification based on abundance of reads per transcript. Unlike the genome, the transcriptome varies across tissues and cell types, and changes dynamically during development and in response to external stimuli. Common tissue sources for transcriptomics include blood with or without cell sorting; bronchial epithelium, smooth muscle or sputum cells for asthma; nasal epithelium for allergic rhinitis; and skin for atopic dermatitis. Different cell types may feature different associations, and this provides insight into how various genes contribute to the many manifestations of allergy.

Recent studies of allergy have identified, across multiple tissue types, differential expression of genes involved in innate and adaptive immunity, inflammatory and repair responses and epithelial integrity. Cytokines (T2-related and others), chemokines and their receptors, host defence proteins (defensins), protease inhibitors (SERPINs) and other multifunctional regulatory proteins (S100 family) are differentially expressed in allergic diseases. SERPINs control various immune and inflammatory processes, for instance by inhibiting neutrophil proteases (elastase, cathepsin G) and fibrinolytic enzymes (plasminogen activators). S100 proteins commonly serve as damage-associated molecular patterns, signals of cell stress or injury. As such, both SERPIN and S100 family proteins probably represent downstream sequelae of the immune-inflammatory responses typically seen in asthma and allergic diseases. Multiple studies have identified such changes for atopic dermatitis, in both lesional and non-lesional skin samples [39–42]; and in airway epithelial or sputum samples of asthma [43–47]. For asthma, further analyses have linked certain transcriptomic profiles to inflammatory subtypes of asthma: eosinophilic or T2-driven airway inflammation has been associated with elevated airway expression of periostin (*POSTN*), *CLCA1*, *SERPINB2*, *CLC*, *CPA3* and *DNASE1L3* [43, 44, 46]; while neutrophilic or Th17-linked inflammation has instead been linked to expression of *IL1B*, *ALPL*, *DEFB4B*, *CXCR2* and other chemokines [43, 47], an expression profile that bears some similarities with psoriatic skin lesions [47]. Differences in gene expression across inflammatory phenotypes are also reflected in blood and sputum transcriptomics [47], and show some promise in being exploitable as putative biomarkers for disease subtypes [43]. Furthermore, there is evidence that T1 and Th17/ILC3 pathways act in partial opposition to each other, and while T2-mediated eosinophilic inflammation is responsive to steroid treatment [44, 48], it may also lead to enhanced Th17 activity and subsequent risk of neutrophilic inflammation [49]. Finally, although it may be enticing to describe T2-mediated eosinophilic inflammation as "allergic", and T1 or Th17-mediated neutrophilic inflammation as "non-allergic", other inflammatory profiles (paucigranulocytic, mixed) also exist, thus complicating the narrative. Nonetheless, associating gene expression profiles with specific inflammatory phenotypes may provide the next step towards improving precision in managing asthma and allergic disease.

It is notable that few of the aforementioned differentially expressed genes were identified as genome-wide significant loci in previous GWAS for asthma. As discussed earlier, this is probably the result of differential gene expression being indicative of inflammatory pathology downstream of the genetics. This is supported by the observation that several expression quantitative trait loci (eQTLs) are located around T2-related loci (*IL4R*, *TSLP*, *IL13*), and that unsupervised gene module analysis of airway transcriptomics has revealed consolidation of certain expressed genes into T1-driven *versus* T2-driven modules [46]. eQTL analyses are similar to GWAS, in that eQTLs are essentially SNPs with genome-wide significant effects on expression of nearby genes, for instance by altering the regulatory region of those genes (cis-eQTLs) [50], or altering a transcription factor for a distant gene (trans-eQTLs). Nowadays it is uncommon to see GWAS without an accompanying eQTL analysis in related tissue types. Significant loci from asthma and allergy GWAS that overlap with eQTLs in specific tissue types (usually whole blood) are shown in supplementary table S1. A limitation of whole blood eQTL analyses is that it is not clear whether the eQTL is active across all blood cell types, or only within specific blood or immune cells.

More recently, single-cell transcriptomics have come to the fore. Scientists can now isolate single cells (*e.g.* micromanipulation with capillary pipettes, flow-activated cell sorting, microfluidics) [51], then investigate transcriptional differences between individual cells within a sample, rather than assuming homogeneous expression and averaging transcription across the sample. Single-cell (sc)RNA-seq presents new opportunities to explore the inner workings of the human immune system, whether it be exploring trajectories of certain types of immune cells ordered by pseudotime, mapping immune cell lineages or investigating B- and T-cell repertoires [52–54]. Of particular future interest is the potential harnessing of scRNA-seq to identify drug or vaccine targets for modifying B- and T-cell responses [53]. Most recently, using scRNA-seq, CROOTE *et al.* [55] identified that certain IgE antibodies of peanut-allergic individuals converged upon identical gene rearrangements. CHIANG *et al.* [56] identified that a subset of Th2 cells (Th2$^+$) in peanut-allergic individuals demonstrated functions beyond IgE isotype switching, such as expression of cytokines that contributed to local tissue inflammation (IL-3, colony-stimulating factor (CSF) 2), as well as resistance to attempted suppression by regulatory (Treg) T-cells. Widespread adoption of scRNA-seq is currently limited by high cost, high computational demand of data processing and inherent challenges in subsequent statistical analyses, specifically in relation to sample normalisation, batch effects and other sources of bias [51, 54]. However, it is anticipated that these obstacles will be gradually resolved with time. There is ongoing development of statistical and systems-based approaches designed to deal with such data, especially measures of trajectory and pseudotime [57].

Epigenomics

The epigenome is the set of heritable biochemical modifications that change gene expression, but are not coded in the DNA sequence. Epigenetics functions as a bridge between genome and transcriptome, providing mechanisms by which the micro- or macroenvironment can influence gene expression within each cell, and by which transgeneration inheritance can occur after initial exposure to an epigenome-modifying environment [58, 59]. Epigenetic signals include 1) DNA methylation at CpG islands, which silences expression of adjacent genes; 2) histone modifications (acetylation, methylation and others), whose effects vary depending on type and position of modification; and 3) noncoding RNA such as microRNA (miRNA), which can silence genes by binding or degrading complementary mRNA [60]. Together, these epigenetic markers cause changes in accessibility of a local DNA segment to transcription or regulatory factors.

Low- and high-throughput detection methods exist for each type of epigenetic signal. Methylation-sensitive restriction fingerprinting and microarrays for detecting 5-methylcytosine have been used to describe the DNA "methylome". Genome-wide histone modifications can be detected using chromatin immunoprecipitation (ChIP). Next-generation sequencing options also exist (miRNA-seq, DNase-seq, formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq, ChIP-seq, 3C-seq), which function by isolating DNA fragments that are accessible or inaccessible to a factor of interest, and sequencing those fragments to determine their identity [9]. Epigenome-wide association studies can then be performed to identify epigenetic features for a given trait or disease. Finally, like the transcriptome, the epigenome is responsive to external stimuli and varies across cell types, and most epigenomic studies of allergy have so far examined blood, skin or airway samples.

There is evidence that development and maturation of T-cell lineages is partly determined by epigenetic changes [58]. Th2 differentiation is driven by *STAT6* and *GATA3*, resulting in epigenetic changes (DNA methylation, histone acetylation) that induce Th2-related (*IL4/IL13*) and suppress Th1-related (*TBET*, *IFNG*, IL-12/STAT4 pathway) expression; conversely, Th1 differentiation is driven by *STAT4* and *TBET* to elicit the opposite epigenetic changes; finally, Treg differentiation is driven by *STAT5*, with associated epigenetic changes in *FOXP3* and the *IL10* locus [58, 60]. Given the role of epigenetics in T-cell

development, it is plausible that allergic disease may be linked to altered epigenetics affecting this process. Epigenetic signals have been observed across multiple tissue types in allergy. Changes to DNA methylation have been noted in loci related to Th2 function and T-cell development (*IL4R*, *TSLP*, *IFNG*, *FOXP3*, *STAT5A*) [59, 61–65], while other significant loci control antigen presentation, eosinophil activity, lipid metabolism and mitochondrial function [66, 67]. The relationship between histone modifications and allergy or asthma is less clear. Some studies have shown changes to global histone acetylation with reduced deacetylating-to-acetylating (HDAC-to-HAT) activity in asthmatic lungs compared to normal [68–70]; while others suggest that HDAC inhibition can improve the suppressive function of Tregs [71]. Similarly, certain miRNAs are known to influence allergy risk. For example, OKOYE *et al.* [72] observed that miR-155 and miR-146 may be critical in determining T-cell differentiation towards Th2 *versus* Th1/Th17. Other relevant miRNAs are reviewed elsewhere [73, 74]. There remain too many knowledge gaps to allow us to fully use epigenetics to our advantage in managing asthma and allergy. However, investigation of the full compendium of miRNA species is progressing rapidly, and may lead to new targeted therapeutics in the future.

An important aspect of epigenetics is the link to environmental exposures. Because the development of the immune system begins *in utero* and continues through infancy, environmental modifiers of epigenetic signals may have a stronger impact earlier in life. Experimental and observational studies show that maternal exposures during pregnancy and exposures during early childhood can modify the child's epigenome. These exposures include changes to diet, macro- and micro-nutrition, farm environments, infections and microbes, animals, allergens, medications, pollutants, tobacco smoke and even maternal stress [60, 75, 76]. In particular, folate and vitamin B12 are methyl donors that have a global impact on DNA methylation [60]. Finally, genome associations have been identified for methylation patterns as quantitative traits (meQTLs). These include the *ORMDL3/GSDMB* locus, where a SNP behaves as both an eQTL and a meQTL [77], and others [66, 78, 79]. All these findings illustrate that certain perinatal exposures can act through genetics and epigenetics to influence disease risk.

The microbiome

The microbiota is the community of microbes, including commensals and pathogens that reside within a host or environment, while the microbiome is the genomic content that represents the microbiota. The "microbiota hypothesis", a modern reiteration of the hygiene hypothesis, suggests that perinatal microbial exposure is vital to proper development of immune functions, especially of tolerance [80–82]. Microbial exposures may modify allergy susceptibility by initiating different trajectories of immune development and function [75]. Epigenetic changes may also be involved in this process, although the exact nature of these changes remains unclear.

The primary interfaces for host–microbe interactions are the epithelial surfaces exposed to the external environment, in the skin and respiratory and gastrointestinal tracts, so most studies on allergy microbiomes involve sampling at one of these sites directly (biopsy or surface samples) or indirectly (faecal or sputum samples). The gut is home to gut-associated lymphoid tissue, and its microbiome can influence disease at other mucosal surfaces, such as the respiratory tract [83, 84]. The respiratory microbiome may exert a direct influence on local inflammatory processes leading to asthma development [85]. The environmental microbiome may drive restructuring of host microbiomes, or modify allergy risk by other means; this may be particularly relevant in relation to the protective effect of farming environments [75]. Description of the microbiome relies mostly on quantification of DNA sequences encoding the 16S ribosomal RNA (rRNA) gene, which is common to all bacteria but contains variable regions used to differentiate taxa. The gene sequence is amplified using PCR and then examined using gel electrophoresis, terminal RFLP, microarrays or sequencing. Recently, there has been a transition to deep metagenomic sequencing, which captures the genomes of all organisms present in a sample, not just the 16S rRNA gene, and can be used to infer both taxonomic composition and function of the microbial community.

Microbiome studies are complicated by the fact that host microbiomes can change with age, season, time of day, site sampled on the host's body and geography [86, 87]. However, a number of consistent findings have been established for asthma and allergy. Features of the gut microbiome associated with allergy include early-life reduction in microbial diversity; reduced populations of Bifidobacteria, Lactobacilli and Bacteroidetes; and increased coliforms and specific Firmicutes (Staphylococci, Enterococci) [83, 84, 88]. Reversing the above changes, for instance by oral administration of certain *Lactobacillus* and *Bifidobacterium* species, may offer some protection against both the initial development of allergy and further exacerbations of atopic disease [80]. Within the airway microbiome, asthma development, symptoms and exacerbation have all been associated with increased Proteobacteria populations (especially *Haemophilus*, *Moraxella*, *Streptococcus* and *Neisseria* spp.), and reduced Bacteroidetes and Fusobacteria commensals [80, 81, 84, 85, 89]. Remarkably, these associations begin during infancy: the detection of

asthma-related bacteria in the first few months of life has been associated with developing allergic asthma by primary school age [81, 83]. Although it is unclear whether microbial changes represent a cause or effect of underlying immune dysfunction, there is evidence of altered gut and airway microbial communities preceding allergic sensitisation [85, 90, 91]. Ultimately, these findings suggest two independent processes at work: microbiota, especially of the gut, exerting systemic effects on immune maturation; and microbiota causing local inflammatory processes at the sites they inhabit, including those associated with asthma in the respiratory tract.

Other recent studies have uncovered the potential role for nonbacterial microbes, including viruses such as human rhinovirus and respiratory syncytial virus, in causing early childhood wheeze and bronchiolitis that often precedes full-blown asthma [75, 85, 92–97]. There is evidence for the role of rhinovirus (RV), specifically RV-C, in causing severe respiratory illnesses that are associated with increased asthma risk later in life. This is further supported by evidence that a genetic locus significant for childhood asthma, CDHR3, modifies the binding and replication of RV-C, and hence infection susceptibility. The pathophysiology behind the viral associations may be related to chronic airway injury due to recurrent infection, possibly interacting synergistically with allergic mechanisms, to elicit and maintain sustained inflammation. Microbe-specific systems such as the virome and the (fungal) mycobiome may also be helpful towards understanding asthma pathogenesis. Microbiome modification and control of respiratory infection risk (e.g. through vaccines or pre/probiotic supplementation) are possible avenues for future investigation.

The exposome and environmental exposures

Researchers have frequently explored the relationship between environmental exposures and disease. The "exposome" builds on this idea by encapsulating all environmental exposures that contribute to human health and disease. The environmental microbiome, for instance, is just one type of exposure; the host microbiome itself can be considered an exposure when describing microbes residing on the skin, or on luminal surfaces of hollow viscera exposed to the external environment. It is difficult to measure all exposures, let alone on a high-throughput scale, and there are other challenges related to correlation, confounding and interaction amongst different exposures [98]. Instead, most studies have so far quantified a limited set of relevant exposures via questionnaires and environmental sampling. NORTH et al. [99] is one of the first studies to adopt an exposomic approach to examine multiple types of exposures simultaneously, in their search for associations with childhood wheeze.

The environment can contribute to asthma and allergy pathogenesis in many ways. As mentioned in the epigenomics section, these include mechanisms acting through diet and nutrition, exposures to pets and animals, allergens, pollution, tobacco smoke and other chemical exposures. For some of these, it is possible to measure and perform high-throughput analyses on proteomic and metabolomic data. Diet is one example: a protective effect against allergy has been reported for polyunsaturated fatty acids (PUFAs) found in fish oil, and for their metabolites [100]. Higher proportions of certain very-long-chain PUFAs in plasma during childhood has been linked to reduced allergic disease in later adolescence [101]. There has been slow adoption of omics-level analyses in food [102], and it remains controversial whether food and dietary supplements have any impact on allergy or asthma risk (given the innumerable potential confounders). However, in the future, it may be possible to scan the contents of an individual's diet in a high-throughput manner, construct a "foodome" (combining lipidomics and metabolomics), and search for de novo associations with disease. Airborne pollutants may also be explored in a similar manner.

Environmental allergens can themselves be investigated by multiple omic approaches, in relation to quantity of exposure, geography of exposure, and allergenicity of protein structures. For instance, studies have identified that low environmental load of allergen can be a risk factor for disease [103, 104]. Timing and route of allergen exposure may also be relevant: early introduction of solids, including peanuts, may be protective, but only within a specific time window [105]. Furthermore, early exposure to peanut allergen through the skin may promote sensitisation, while exposure through the gut may promote tolerance [106]. Other studies have overlaid geographical maps of exposure with maps of disease, as has been done for traffic-related air pollution and asthma [107]. Finally, it is still not clear why allergens behave as allergens, or more specifically whether there is anything inherent in the molecular structure of putative allergens that confer allergenicity. The term "allergome" is typically used to describe the proteomics-based discovery of allergenic protein structures within individual allergens (see the section on proteomics for discussion).

Occupational and chemical exposures comprise a less common, but well-known source of irritants and allergens that cause asthmatic disease. In occupational health, the term reactive airways dysfunction syndrome is often used to refer to bronchial reactivity without an initial latency or sensitisation period [108]. Examples of culprit chemicals include isocyanates, acid anhydrides, azodicarbonamide, dyes, enzymes and

metals [109]. Potential mechanisms of disease are highly variable, and may involve type I (allergic) hypersensitivity *versus* non-allergic/irritant, or type IV (T-cell-mediated) mechanisms; "inducers" *versus* "inciters"; dermal *versus* respiratory sensitisers; and low molecular weight (hapten-like) *versus* high molecular weight allergens [108]. It is likely that many culprits drive disease through mixed mechanisms. An important note here is that a current lack of evidence for IgE-mediated mechanisms with a particular chemical trigger does not rule out allergy to that chemical as a cause; limitations still currently exist in the engineering of appropriate detection methods for chemical-specific IgE [110].

As alluded to previously, environmental exposures can act through interactions with host microbiome to modify disease risk [75, 83, 84]. Maternal and perinatal exposure to rural environments confers some protection, possibly due to contact with microbial products such as lipopolysaccharide, greater diversity in microbial exposure or environmental modification of host microbiota. Caesarean deliveries and perinatal use of antibiotics may increase risk for allergy, possibly by disrupting neonatal microbial colonisation. The protective effect of oral probiotics with Lactobacilli and Bifidobacteria spp. has been reported, as noted previously, and they may also provide cross-organ protection, reducing the incidence and severity of respiratory infections [84]. The use of dietary fibre in prebiotics, with subsequent fermentation into short-chain fatty acids, may protect from allergy *via* Toll-like receptor and G protein-coupled receptor signalling or epigenetic modifications [80, 81]. Vitamin D has potential immune and microbiome-modifying effects, and vitamin D deficiency is a suspected risk factor for allergy [111, 112]. Breastmilk contains immunoactive molecules and may alter gut microbiota composition [80]. Altogether, these findings offer a glimpse into how multiple environmental exposures may interact in a complex fashion to elicit disease.

Proteomics, metabolomics and lipidomics

The proteome is the repertoire of proteins produced by cells or tissues, reflecting the molecular effectors and metabolic consequences of cell function. Common proteomic technologies can be grouped into antibody-based (ELISA), peak-profiling mass spectrometry (MS)-based ("fingerprinting"), gel-MS based (1D/2DG, 2D-DIGE) and liquid chromatography-MS based methods [113]. The general approach is to perform coarse separation of digested proteins into "bands" or "spots", and then further investigate each spot by MS. The MS steps are often conducted in tandem (MS/MS) to achieve higher resolution. The information gained from MS can then be used to identify the peptide or construct its amino acid sequence. Sources of proteomic samples include sites of pathology such as the airway, in the form of cellular or fluid content from bronchoalveolar lavage, induced sputum, biopsies or *in vitro* cell cultures; or it may involve the usual blood or urine sample [114]. An accessible type of specimen unique to asthma research is exhaled breath condensate (EBC), which provides information on volatile compounds released from the airway.

In relation to asthma and allergy research, proteomic changes often depict nonspecific pathology, as in a general elevated inflammatory state, as well as underlying pathological mechanisms. Therefore, recent findings in proteomics mostly mirror transcriptomic changes, in that they reflect altered functions in immunity, inflammation and antiprotease activity: affected proteins include defensins, α_1 antitrypsin, α_2 macroglobulin, SERPINs, S100-family proteins, apolipoproteins and complement proteins [113–116]. Of particular note is a recent study by SCHOFIELD *et al.* [117], which combined sputum transcriptomic and proteomics with airway histology and clinical features. They found that eosinophilic phenotypes were associated with increased blood periostin and sputum haptoglobin, while neutrophilic phenotypes were associated with increased S100A9 and MMP9. Interestingly, few recent studies have identified proteome-wide significant changes to T2-related cytokines, although associations have been found within low-throughput *in vivo* studies in the past [118]. It is plausible that, being upstream of signalling cascades, these T2 cytokines are less apparent in proteome-wide analyses, where significant findings tend to be more dominated by downstream proteins that have been amplified *via* the cascades.

Another important contribution of proteomics to allergy research is allergen detection and discovery [119]. Studies have investigated a compendium of epitopes for aeroallergens such as house dust mite [120–122] and plant pollen [123–125], and for food allergens in seafood and processed foods [119]; these have served both to confirm existing epitopes and to identify new ones. Findings from these studies can be applied to nonclinical settings, such as food processing and safety [119].

Metabolomics is the systems-level study of metabolites, the nonpeptide macromolecules representing the substrates and end-products of cellular activity. The two main technologies of measurement used in metabolomics are nuclear magnetic resonance, which provides a spectral fingerprint of a system's metabolite constituents, and MS. Like proteomics, most metabolomic studies focus on samples of blood serum, EBC and urine from asthmatic patients [126–128]. Lipidomics is a subset of metabolomics specifically dealing with lipid molecules, and lipidomic studies have shown that allergic disease is typically

associated with elevation of arachidonic acid metabolites belonging to the lipoxygenase pathway, such as leukotrienes [129, 130]. Metabolomic associations with asthma involve immune and inflammatory functions, oxidative stress and hypoxia, cellular energy homeostasis and lipid metabolism pathways [127]. These associations seem to reflect general biological stress or inflammatory pathology, rather than specificity for allergy or asthma. However, predictive and discrimination models based on metabolomic findings have shown some promise [127]. Ultimately, as was the case for transcriptomics, proteomics and metabolomics are being used to identify potential biomarkers to screen for asthma and stratify into asthma subphenotypes. Additionally, in line with developments towards integrating multiple omics: proteomic, metabolomic and lipidomic methods may be applied not just to host samples, but also to environmental samples.

The phenome and physiome

Phenomics is a broad term encompassing all physical or biochemical traits (phenotypes), observable in cells or individuals, that reflect states of disease or health ("physiome"). In the case of allergy and asthma, possible phenotypes include cell types based on morphology and response (immunophenotyping); clinical biomarkers, such as antibody assays and cell counts; and the extensive physical manifestations of disease, embodied in clinical history, symptoms and signs and investigation results. These traits may be quantified and described in detail, although not necessarily using high-throughput technologies. Phenotypic traits of interest may also include nondisease states, such as clinical remission, and traits that vary with age. Integration with other omics-based datasets, and incorporation of time scales into analyses, may yield further insight as to how such resilience against asthmatic disease is conferred.

Phenome-wide association analyses, where large sets of traits are screened for enrichment of allergy-related genetic loci [131, 132], have been performed in the past, but have yet to gain widespread popularity. Phenomes and phenotypes can also be analysed by machine learning, whether it be comparison of known phenotypes (via supervised classification) or construction of new phenotypes from omic or non-omic data (via unsupervised cluster analysis). This is discussed in further detail later.

The "immunome" is a subset of the physiome that is highly relevant to allergy, and where high-throughput technologies play a major role. Immunomics broadly describes the systemic quantification of immune function by examining immune cell populations and expression of immune mediators. It may use immunoglobulin [133–135] and cytokine (proteomic or transcriptomic) arrays [136, 137] to quantify immune responses such as sensitisation, in vivo or in vitro. It can also involve leukocyte immunophenotyping and high-dimensional or mass cytometry [138–141]. The immunome is complex and varies dramatically by sampled immune cell type, tissue or organ, age and timing of sampling, especially before and after sensitisation. Using these types of data, a number of recent studies have begun exploring the "core immune signatures" of newborn infants and adults alike [142–144]. Although it is well known that allergy is a T2-driven phenomenon, it is still not clear how all the components interact to generate disease, nor is it clear how heterogeneity in disease or health is explained by immunome heterogeneity. Furthermore, the nonallergic contributions of asthma are not well understood, and it is unclear how T2 and non-T2 (particularly T1 and Th17/ILC3) mechanisms interact to generate the spectrum of disease. Future studies may be able to shed light on this.

One major aspect of human physiology that has a known impact on asthma risk and progression is sex. There are clear differences between males and females in terms of the development and physiology of the immune [145] and respiratory systems [146]. During early childhood, males have a higher incidence of asthma or wheezing illness than females; this switches during and after adolescence [147]. Pregnancy in females can often exacerbate asthma symptoms, with partial resolution postpartum. It is likely that hormonal changes during these life phases influence these changes in disease risk. Current evidence suggests that ovarian hormones and oestrogen-mediated signalling promotes both Th2-related and Th17-related inflammation, while testosterone is protective against Th2-mediated inflammation [147]. While existing studies already account for sex and gender differences as an important covariate or cofactor, future studies may further explore sex interactions with immunomics, in particular, investigating how sex hormones modify the function of both adaptive and innate immune cells.

Integration of omics data

Following our overview of the omics, we now discuss common techniques used to integrate and interpret omics data in allergy and asthma research.

Exploring intra- and inter-omic relations

To understand disease pathogenesis, it is natural to compare findings across different omics, and construct a multi-omic model of pathophysiology that links these various elements together. This may be a simple sequential model of causality, or a complex network of interacting components. Many studies on omic

associations with allergy and asthma also search for inter- and intra-omic relationships. Relationships can take the form of direct associations, where one entity behaves as a trait for another, or an interactive effect between two entities in relation to a third entity as the trait of interest. The study of these relationships is the crux of modern systems biology.

Genomics, being the most studied system in allergy and asthma, features extensively in intra-omic and inter-omic analyses. GWAS can be found not only for clinical phenotypes (*e.g.* presence of allergic disease) as traits, but also for expression of transcripts (eQTL analyses), epigenetic markers (meQTLs) and intermediate phenotypes such as microbial exposures and immunomes. Recently, there has been a concerted move towards integrative genomics and genetic effects on gene expression are a pervasive component of modern association studies, in the form of mandatory genome-wide eQTL analyses or targeted measurements of gene transcripts [74]. Also coming into vogue is the use of Mendelian randomisation, a technique which uses genomic information as instruments to infer causal links between one trait or phenomenon and another, based on the assumption that allelic genotypes are randomly assigned as they are passed from parent to offspring [148]. The traits being linked may themselves be related to gene loci or expressed genes [149].

Analyses for interactive effects with other omics also feature heavily in allergy genomics. It is unlikely that genetic and environmental factors act independently in conferring risk, so modern genomic studies often include interaction terms with exposure variables. Scientists have explored interaction effects on asthma susceptibility between genetics and exposures such as air pollution and tobacco smoke [150, 151]. Another example is the impact of allergen exposure and genetics on immune cell gene expression [152]. Interaction analyses also extend beyond environmental effects. Gene–ethnicity interaction has been investigated *via* admixture mapping [10]. Genetic–epigenetic interactions have been reported; some genome-wide significant loci (*e.g. IL4R*) may interact with nearby epigenetic signals to alter disease risk [65]. While investigation of gene–gene interaction (epistasis) is of intense interest, the overwhelming number of active genes in the human genome means that such analyses have a large statistical burden and hence remain difficult. Therefore, gene–gene interaction studies are so far limited to a few selected genes or SNPs. Polygenic risk scores tend to employ additive linear models that do not reflect epistatic effects, and it remains a challenge to integrate these in an accurate manner. Finally, interactive effects may be explored by means beyond using interaction terms in regression models, for example, eQTL-weighted GWAS have been reported [153].

Given the strong links between environmental factors and asthma, interactions with environment exposures have been explored to a degree. Importantly, prominent gene–environmental interactions have been observed with glutathione S-transferase variants impacting on susceptibility to environmental sources of oxidative stress, such as air pollutants [154], passive exposure to tobacco smoke [155] and isocyanate [156, 157], as well as subsequent asthmatic disease [154, 158]. Similar interactions may exist for respiratory infections, given that infection and the subsequent immune-inflammatory reaction are also sources of oxidative stress. In addition, microbial and pathogen exposures have been linked to differential gene expression, for instance, viral infections are associated with changes to airway epithelial transcriptomics in asthma [159, 160]. Unsurprisingly, the exposome and microbiome have also been linked to epigenetic changes, and the various exposures are intricately entwined in complex interactions. For instance, a recent study has looked at the interaction between air pollution and the allergenicity of ragweed pollen [161]. Another recent study has identified that maternal phthalate exposure may promote allergy in subsequent generations *via* epigenetics [162]. Other examples concerning environmental interactions with diet and microbiome have already been discussed.

Finally, a common application of integrative omics is the use of gene ontology analysis to annotate discovered genes from genomic, transcriptomic or epigenomic analyses [163]. This makes use of a pre-curated database of functional annotations for known genes, based on existing literature, to segregate discovered genes into groups or pathways with shared functions. An example is the Gene Ontology Consortium [164]. These databases of functional annotations convey phenomic information, where cell phenotypes, functions and behaviours are organised into discrete categories. In doing so, one aims to condense diverse genome-wide findings into concise summaries of biological function that may be easier to interpret when building a conceptual model of pathophysiology. Similar annotation analyses exist for proteomics [165, 166]. A limitation of such techniques is that the annotations may not always be certain, reliable or up-to-date, and can often be vague or uninformative.

Inter- and intra-omic relationships may be explored either by low-throughput pairings, or by high-throughput assessment of larger networks [167, 168]. However, especially with the latter, it may be difficult to account for noncausal correlations or confounders. For example, despite the hygiene hypothesis, low socioeconomic status and impoverished environments remain risk factors for the

development and severity of asthma [84]. This may be due to confounding factors that coexist with poverty, including urbanised environments, exposure to allergen and pollutants, dietary intake and access to healthcare. There is no doubt that modifiers of allergy risk may co-occur, but whether this represents a causal link is another matter. Methods such as Mendelian randomisation (described earlier) may be used to disentangle this, but one must be wary of violating the numerous assumptions that underlie Mendelian randomisation. In addition, given the high dimensionality of inter- and intra-omic analyses, dimension reduction and machine learning may be used instead to identify potentially robust signals of relevance to pathogenesis.

Machine learning, dimension reduction and clustering

Machine learning is a set of methods that use computing to learn and formulate solutions from supplied data, with or without explicit human input. It is already in common use with various biomedical and ecological applications [169–171]; however, it is particularly useful when dealing with complex, high-throughput and multidimensional data, especially in cases where pre-existing human knowledge may be unavailable or insufficient to decipher the data. Machine learning approaches typically involve iteration, where an algorithm repeatedly refines a model based on observed data until a metric of model quality (*e.g.* objective/cost/loss function) satisfies a particular threshold. Applications of machine learning in biomedicine typically involve the exploration of data structure, or generating predictive or explanatory models of biological systems.

Cluster analysis and classification are methods used to subset data samples or individuals into different groups or categories, thus giving a summary of data structure. Such methods typically employ machine learning at the most fundamental level: for instance, hierarchical clustering is an iterative process where the "objective function" would be the minimisation of within-cluster similarity and/or maximisation of between-cluster dissimilarity. There is usually a subtle distinction between clustering and classification: cluster analysis is a data-driven approach, where omic data is used to generate clusters in an unsupervised fashion. The clusters can then be interpreted for hypothesis generation and testing. Conversely, classification is a hypothesis-driven approach: known phenotypes or pre-curated categories are used to determine a model of classification based on training data, which can then be applied to other datasets, or examined to look for further biological associations (figure 3).

A drawback of clustering and classification (as for other applications of machine learning) is that there is little consensus or standardisation of optimal methods, although there are certainly favoured approaches for each problem. In addition, they may be intimidating for the regular clinician or biologist to adopt, and choice of method often depends on a specialist understanding of nuances in the data. As an example: when performing cluster analysis, many decisions need to be made prior to and during the procedure. This includes how to deal with missing data; select the variables or "features" for clustering; scale or normalise features; choose the algorithm to do the actual clustering; pick the number of clusters; control for overfitting; and validate or replicate results [172]. It is not necessarily clear what the best choice is for any of these decision points.

In exploring the correlation structure and confounders in a dataset, principal components analysis or similar methods can be used to transform the dataset into uncorrelated variables or "principal components". In doing so, we can observe which of the original variables describe similar information (*i.e.* are highly correlated with each other), and by plotting principal components, we can visualise the data in a way that maximises variability between samples or variables. By condensing our data to a limited selection of principal components, we can reduce the number of dimensions and simplify the input features for subsequent clustering or classification [173]. Feature selection can be limited to a single omic entity, or cover multiple omics simultaneously, depending on the question asked.

Cluster analysis involves separating samples in a dataset into discrete groups (clusters) based on what can be learnt from data structure, without specifying training examples for each group [172, 174]. Its objective is to minimise intragroup differences and maximise intergroup differences. Measures of difference or dissimilarity may be distance- or correlation-based. Common clustering techniques include hierarchical clustering, medoid-based methods, and latent variable modelling. Cluster analysis allows the identification of homogeneous groups within a heterogeneous dataset, and simplify analyses to comparisons between clusters rather than across entire cohorts. Clustering can also expose confounders without explicit adjustment for correlation, especially if clustering is "guided" by cosegregating omic variables. Using molecular omic-based features, cluster analysis may allow us to determine endotypes (subtypes of disease or health states) by common biomolecular interactions and pathophysiology [175]. These can be compared with known phenotypes to explore how variation in pathophysiological mechanisms are linked to variation in disease manifestations. In addition, cluster analysis can be applied to phenome data to deal with
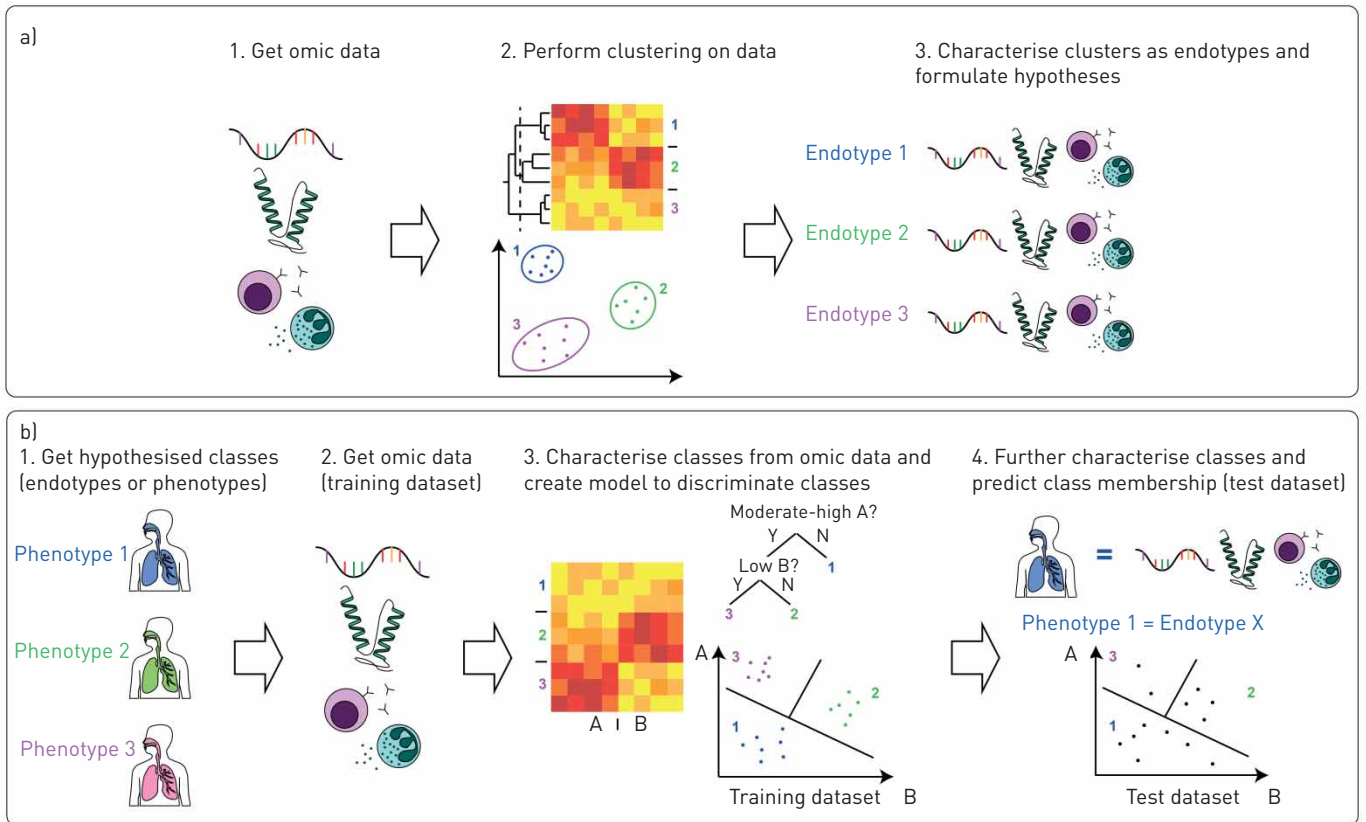
FIGURE 3 Data-driven *versus* hypothesis-driven machine learning for integration of omic data. a) Data-driven (unsupervised) cluster analysis used to generate *de novo* groupings, reflective of shared pathophysiology ("endotypes"); b) hypothesis-driven (supervised) classification to compare known phenotypes or endotypes, and to allow prediction of phenotype/endotype membership for additional samples. Diagrams are for illustrative purposes only and do not convey real data.

heterogeneity in phenotypes. Using "cleaner" subphenotypes for association analyses may improve the power and specificity of subsequent findings.

Classification methods determine a statistical model or decision-making algorithm that allocates individuals of a training dataset into known groups (classes) [172]. The learnt model or algorithm can then be applied to other test datasets for classification into classes. Methods include regression analysis, discriminant analysis, support vector machines and partitioning or decision trees. The objective of classification varies with the method, but mainly involves achieving the "best fit", minimising differences between predicted and actual class allocation for the training dataset, without compromising generalisability to external datasets. Classification can be used to design diagnostic or risk stratification algorithms from an omic dataset. Each sample is labelled as one of a predefined set of phenotypes (*e.g.* allergic *versus* nonallergic asthma, eosinophilic *versus* neutrophilic, severe *versus* nonsevere), then the algorithm seeks biomolecular or clinicophysiological features that best define the phenotype [176, 177]. In the absence of predefined phenotypes, clustering and classification may be combined: clusters are generated based on a training dataset, then a classifier is devised which can classify test datasets into the discovered clusters.

Both cluster analysis and classification have been used extensively in asthma research. Major findings from such analyses include the discovery and characterisation of different subsets of childhood and adult asthma. Childhood wheeze has been categorised, by both traditional and machine-learning approaches, into persistent atopic wheeze of early onset, transient remitting viral wheeze, and a mixed atopic/nonatopic phenotype of variable onset [178–180]. Atopic wheeze appears to be characterised by Th2 activation, early sensitisation to allergens, greater severity of respiratory disease, greater likelihood of persistence to full-fledged allergic asthma and concurrence of other atopic diseases. In terms of adult asthma, there are subtypes based on lung function [181], as well as atopic, nonatopic, mixed and other phenotypes [175]. Eosinophilic, neutrophilic and paucigranulocytic airway inflammation can be distinguished from sputum samples, and accompanying transcriptomic, proteomic and immunomic data can provide some insight

into underlying pathophysiology for each phenotype [43, 176, 182–184]. Neutrophilic, Th1/Th17-dominant, and steroid-resistant asthma tend to co-occur, suggesting a common endotype. Asthma, COPD and mixed asthma/COPD phenotypes have been explored [185]. Other studies have looked at allergy phenotypes related to degree and pattern of allergic sensitisation (mono- *versus* poly-sensitised; early- *versus* late-sensitised) [186, 187].

Clustering can be applied to other omic data, other than phenotype data. In TEO *et al.* [85], hierarchical clustering was used to generate the microbiome profile groups which categorised the infant nasopharyngeal microbiome into discrete clusters based on microbial abundance. This facilitated simpler analysis and interpretation of otherwise complex data.

Some researchers have identified that membership within asthma clusters or subphenotypes change or transition with age [188, 189]. This latter point highlights an ongoing challenge of subphenotyping asthma: the fact that these phenotypes or clusters are inherently unstable, and may change with age complicates *post hoc* analyses. To address this, clustering can be applied in a time-dependent manner: several research groups have used techniques (*e.g.* latent transition analysis) that leverage longitudinal data to model transition probabilities between clusters at different time points [189–191]. Such methods reveal which asthma phenotypes are inherently stable or unstable; at a cursory glance, it appears that early-life atopy tends to correlate with entrenched asthma in later life [191, 192]. Our own laboratory recently employed a method of cluster analysis to derive trajectories representing distinct patterns of evolving composition in the nasopharyngeal microbiome, and subsequently related these to asthma outcomes [193].

Network analysis

Network analysis is the use of networks to model and investigate systems. Networks are represented by graphs consisting of nodes and edges, where nodes represent entities (*e.g.* biomolecule) and the edges between nodes indicate relationships between entities (*e.g.* correlation, transition probability, molecular interaction). Edges can be undirected (symmetrical) or directed (asymmetrical). Many types of network analyses involve use of machine learning to generate a best-fitting network for a given dataset.

Networks are used to discover and visualise how different components in a system relate to each other, whether they be abstract relations or actual molecular interactions. Bayesian network analysis involves probabilistic modelling of a network, where edges are directed and annotated with a transition probability from one node to another. This technique has been used frequently in asthma research, for instance to identify candidate genes or SNPs associated with a bronchodilator response [194]; to quantify interactions between measured pathophysiological variables related to asthma and allergy [195]; and to describe gene regulatory networks using gene expression and GWAS data [196, 197].

Gene co-expression networks can be generated based on correlation between expression levels of different genes. High correlation reflects genes that are co-expressed and hence may be co-regulated or share a common biochemical pathway. Nodes represent genes, while edges represent correlation between them. Furthermore, edges can be weighted by degree of correlation, as in weighted gene co-expression network analysis (WGCNA); and highly connected or proximate subgraphs can be interpreted as gene modules of functional importance. WGCNA has been used to identify co-expression networks underlying helper T-cell responses to house dust mite stimulation [198], transcription networks in whole blood of asthmatics [199], an IgE-signalling gene network associated with blood lipids [168], and co-methylation models that reflect asthma endotypes [77]. MODENA *et al.* [46] identified that adults with severe asthma have differential airway expression of gene modules corresponding to various biological functions, including epithelial growth and repair, T1/T2 inflammation, neuronal and cilia functions. In particular, certain subsets of severe asthmatics exhibited high expression of a T1-associated gene module, featuring core genes *STAT1* and *PARP9* as well as other notable downstream proteins (*e.g.* interferon-γ induced chemokines).

Other applications of network analysis exist. For example, PILLAI *et al.* [200] used bipartite network analysis of cytokine expression to sort patients into distinct endotypes. HINKS *et al.* [195] constructed a network of asthmatic individuals based on similarity in clinicophysiological parameters, then used topological data analysis to assign nodes into clusters.

Finally, the term "network analysis" has been used to describe the application of genomic, transcriptomic or proteomic data to existing networks stored in databases, specifically protein–protein interaction networks, or networks representing biomolecular pathways. This is often done to generate subsets of the original interaction networks, which are then examined for biological interpretation [201, 202]. Network databases concerning other omics may be used to achieve a similar purpose (*e.g.* ingenuity pathway analysis, InnateDB) [198, 203].

Mathematical modelling and prediction

The ultimate goal of integrative analyses is to generate models that reliably explain biological phenomena. At the simplest level, identified omic associations can be used as biomarkers; to generate a model consisting of the strongest biomarkers; and test the model on an external dataset. Many examples of such an approach exist in the literature [128, 204, 205]. At a deeper level, multiple biomarkers (potentially omic-wide) may be aggregated into a risk score, such as a genomic risk score [34]. The classification and network models discussed previously are themselves mathematical models that cover multiple omic domains and are testable on external datasets. In some of these applications, the models represent abstract attributions of risk, and strive to be useful as clinical predictive tools, rather than to be accurate or comprehensive representations of pathophysiology.

Another approach is to observe the consequences of perturbing a system, and infer normal function based on the results [6]. These perturbations may include gene knockouts in animal or cell models [206], neutralising antibodies or receptor antagonists to observe subsequent disruption of function [207, 208], or simple observation of distribution of perturbations among cases *versus* controls in an observational study (*e.g.* GWAS). Perturbations may be deliberate and controlled, targeting single genes or molecules; or they may be randomised and wholesale, in keeping with the "systems" philosophy (*e.g.* random mutagenesis studies in animal models). While such approaches have been employed extensively in biological research, there have been few in-depth explorations of the consequences of perturbation at multi-omic levels, at least in asthma research. Research groups with access to multi-omic datasets may be well positioned to begin exploring such questions.

Based on existing knowledge from biomarker and perturbation studies, it may be possible to generate *in silico* mathematical models to describe a complete biological subsystem in terms of components, interactions and functions; and then describe their perturbation during disease. Modelling biological systems in such a manner is challenging, as there are still many unknowns about its components. However, this has not stopped researchers from trying; for example, HÖFER *et al.* [209] modelled the IL4-dependent activation of GATA3 transcription in T2 development. Multi-scale approaches have been used to describe multiple levels of biological function, from intracellular molecular processes, to cell-to-cell communication, to organ-level function. For instance, LAUZON *et al.* [210] formulated a model of airway hyperresponsiveness that accounted for actin–myosin mechanics, calcium signalling in airway smooth muscle (ASM) regulation; mechanical forces of airway narrowing, and time-dependent distribution of ASM contraction throughout the lung. Such approaches require knowledge of techniques that use differential equations and state diagrams; a review of these approaches is provided elsewhere [211]. However, since such models are usually generated with data from *in vitro* systems or animal models, it remains an ongoing challenge to test their relevance to *in vivo* human systems, and they should therefore be treated with caution [212]. Upcoming projects such as the Human Cell Atlas [213] seek to address some of these challenges, bridging the gap between cell biology and clinical medicine.

Pitfalls and challenges

Many challenges remain for systems biology. There are methodological challenges associated with statistical power, even in large consortia. This is due to the sheer scale of omic data, and the number of possible omic–omic comparisons or interactions. Next-generation technologies are becoming cheaper and more efficient, but the volume of data they generate will continue to pose a statistical challenge. Furthermore, the theory behind statistical and modelling methods still lags behind, and there is currently little consensus on the optimal systems-level pipelines (*e.g.* RNA-seq). Although research groups have recently been paying more attention to measurable environmental exposures in terms of their impact on biological systems [99], the lack of environmental data, and the uncertainty about which exposures actually matter, hinders examination of gene–environment interactions. Finally, even if we have a sufficiently powered sample, there is a so-called Faustian bargain [214], where large sample sizes introduce heterogeneity in cases and controls, thus obscuring findings. There is also a similar problem of the winner's curse [215], where significant results in a ome-wide study tend to exhibit larger effect sizes than they are in reality.

Machine learning has been the go-to tool to handle phenotypic heterogeneity [216]. However, many biologists and clinicians remain sceptical of it, with concerns about its hype or fad-like status, its opaque "black box" nature, and the perceived lack of clear, consistent or immediately-applicable results [217]. MOORE *et al.* [181] were among the first groups to apply unsupervised cluster analysis to an adult asthma cohort and identify distinct clusters. While the clusters themselves proved useful in describing disease risk and severity profiles in the discovery population, subsequent studies attempting to replicate these clusters in other cohorts have had mixed results [218]. Numerous other studies have identified different sets of clusters based on different parameters and populations [219]. Results of machine learning methods may

vary significantly depending on the nature of the input data, in terms of its quality, its relevance to the disease being studied, its depth (resolution of data: categorical *versus* continuous) and breadth (single *versus* multiple biological domains) and its balance (one domain prioritised over another *versus* all treated equally). The variability in research outcomes may suggest to some that machine learning methods are ultimately unreliable, but the field is still growing, and we argue that it simply illustrates the immense complexity of biomedical systems, complexity that will remain impenetrable if we limit ourselves to traditional expert-driven approaches. Unsupervised machine learning can serve as a springboard for future hypotheses: Lazic *et al.* [220] used unsupervised latent class analysis to identify a high-risk multiple-sensitised subgroup, whose pathophysiological origins in early life may be worth exploring in further detail with hypothesis-driven approaches. Ultimately, a balance of human expertise and machine learning will be necessary to make the right decisions about data input and interpretation, and to transform big data into biomedically relevant results.

Systems biology is multidisciplinary, and with this comes another challenge: communication and collaboration between the various disciplines. There is often a conflict of priorities: a clinician might be more interested in diagnosis, treatment and prognosis; an immunologist in the pathophysiology of allergy and asthma; the biostatistician in making sure that the statistics and modelling are sound; and the bioinformatician in generating clean data and writing problem-free code. There may be residual scepticism amongst some biologists or clinicians who perceive systems approaches as "data fishing" [9]. There is some evidence to suggest that multidisciplinary research projects have greater difficulty in getting funded or making a strong scientific impact [221], and this may reflect the challenge of balancing multiple priorities and conveying different perspectives to a broad audience, more so than the actual quality of the writing or research.

Multiple reviews have highlighted the ongoing inaccessibility of systems approaches to many biologists and clinicians, and have recommended the creation of biologist-friendly tools [138, 211, 222]. While this may indeed be helpful for common or simple analyses, there remains an ongoing need for specialist input in developing and using new tools. Tools are only useful if applied correctly, and a research group should not eschew specialist statistics or informatics input, simply to save costs or to keep things simple. Clearly, systems biology is itself very diverse, covering multiple avenues of inquiry. Subspecialties are likely to emerge within the field, each focusing on specific methodologies and their applications. It is likely that there will be a demand for specialists and generalists alike, and the movement of tertiary institutions towards incorporating mathematics, statistics and informatics in undergraduate biomedical courses is certainly a welcome one.

Future directions and concluding statements

The recent developments in systems biology exemplify the global drive towards systems medicine [187, 223], and more broadly, "P4" medicine: predictive, preventative, personalised and participatory [224]. Our ultimate objective is to achieve a critical level of biomedical understanding that permits development of precise and personalised interventions for individual patients. The employment of systems biology in asthma research represents the first step towards achieving this goal. Omics-level research allows us to hone in on the myriad pathological changes that contribute incrementally to disease, and to attempt reversal of these changes by addressing new therapeutic targets (supplementary table S1). Omics–omics integration may enable us to connect these changes and visualise how they operate in each individual patient. Another important contribution of systems biology is the ongoing clarification of the hygiene/microbial hypothesis; the interaction between genetics and environment; and the future possibility of environmental and host microbiota modification to manage or prevent disease. Therapies may be tailored to individual patients depending on their underlying endotype or pattern of pathophysiology (*e.g.* anti-T2 biologic treatment for allergic asthma; combined anti-Th2/Th17 treatment for steroid-resistant disease), though the exact implementation of such precision medicine remains a work in progress.

Worldwide, there has been a push by many groups to implement systems medicine, charting a path from wet lab to dry lab to bedside. Large consortia, such as MeDALL (Mechanisms of the Development of Allergy) in Europe [187] and STELAR (Study Team for Asthma Life Research) from the UK [216], have been established specifically to record and integrate multi-omic data related to allergy and asthma, and conduct well-powered systems-based analyses. Other smaller groups are also involved in similar research *via* frequent cross-collaborations: these include the Childhood Asthma Study (Australia) [85], U-BIOPRED (Unbiased Biomarkers in Prediction of Respiratory Disease Outcomes; European) [225], Childhood Origins of Asthma (USA) [93], Copenhagen Prospective Study on Asthma in Childhood (COPSAC; Denmark) [14], Manchester Asthma and Allergy Study (UK) [186], Severe Asthma Research Program (USA) [182] and others. In the modern age of systems biology, collaboration and data sharing is virtually mandatory when it comes to uncovering complex associations such as gene–environment interactions.

Overall, systems biology has yielded fruitful outcomes in asthma research, and promises to deliver more in the future. At the moment, we are still a far way off from truly personalised medicine – being able to predict with reasonable accuracy the disease or prognosis of an individual based on well-sampled data. However, we can only expect the field to grow exponentially in the years to come.

References

1   Gern JE. The Urban Environment and Childhood Asthma study. *J Allergy Clin Immunol* 2010; 125: 545–549.
2   Bousquet J, Anto J, Auffray C, *et al.* MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy* 2011; 66: 596–604.
3   Eberhardt M, Lai X, Tomar N, *et al.* Third-kind encounters in biomedicine: immunology meets mathematics and informatics to become quantitative and predictive. *Methods Mol Biol* 2016; 1386: 135–179.
4   Kirschner MW. The meaning of systems biology. *Cell* 2005; 121: 503–504.
5   Molinelli EJ, Korkut A, Wang W, *et al.* Perturbation biology: inferring signaling networks in cellular systems. *PLoS Comput Biol* 2013; 9: e1003290.
6   Ge H, Walhout AJ, Vidal M. Integrating "omic" information: a bridge between genomics and systems biology. *Trends Genet* 2003; 19: 551–560.
7   Arazi A, Pendergraft WF 3rd, Ribeiro RM, *et al.* Human systems immunology: hypothesis-based modeling and unbiased data-driven approaches. *Semin Immunol* 2013; 25: 193–200.
8   Dada JO, Mendes P. Multi-scale modelling and simulation in systems biology. *Integr Biol* 2011; 3: 86–96.
9   Bunyavanich S, Schadt EE. Systems biology of asthma and allergic diseases: a multiscale approach. *J Allergy Clin Immunol* 2015; 135: 31–42.
10   Gupta J, Johansson E, Bernstein JA, *et al.* Resolving the etiology of atopic disorders by using genetic analysis of racial ancestry. *J Allergy Clin Immunol* 2016; 138: 676–699.
11   Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev* 2011; 242: 10–30.
12   Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology* 2017; 6: e165.
13   Ortiz RA, Barnes KC. Genetics of allergic diseases. *Immunol Allergy Clin North Am* 2015; 35: 19–44.
14   Bønnelykke K, Sparks R, Waage J, *et al.* Genetics of allergy and allergic sensitization: common variants, rare mutations. *Curr Opin Immunol* 2015; 36: 115–126.
15   Portelli MA, Hodge E, Sayers I. Genetic risk factors for the development of allergic disease identified by genome-wide association. *Clin Exp Allergy* 2015; 45: 21–31.
16   Zhu Z, Lee PH, Chaffin MD, *et al.* A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat Genet* 2018; 50: 857–864.
17   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.
18   Torgerson DG, Ampleford EJ, Chiu GY, *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011; 43: 887–892.
19   Bønnelykke K, Sleiman P, Nielsen K, *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.
20   Imboden M, Bouzigon E, Curjuric I, *et al.* Genome-wide association study of lung function decline in adults with and without asthma. *J Allergy Clin Immunol* 2012; 129: 1218–1228.
21   Yao TC, Du G, Han L, *et al.* Genome-wide association study of lung function phenotypes in a founder population. *J Allergy Clin Immunol* 2014; 133: 248–255.
22   Larkin EK, Hartert TV. Genes associated with RSV lower respiratory tract infection and asthma: the application of genetic epidemiological methods to understand causality. *Future Virol* 2015; 10: 883–897.
23   Pasanen A, Karjalainen MK, Bont L, *et al.* Genome-wide association study of polymorphisms predisposing to bronchiolitis. *Sci Rep* 2017; 7: 41653.
24   Soler Artigas M, Loth DW, Wain LV, *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082–1090.
25   Busse WW, Israel E, Nelson HS, *et al.* Daclizumab improves asthma control in patients with moderate to severe persistent asthma: a randomized, controlled trial. *Am J Respir Crit Care Med* 2008; 178: 1002–1008.
26   Luo J, Liu D, Liu CT. The efficacy and safety of antiinterleukin 13, a monoclonal antibody, in adult patients with asthma: a systematic review and meta-analysis. *Medicine* 2016; 95: e2556.
27   Wenzel S, Wilbraham D, Fuller R, *et al.* Effect of an interleukin-4 variant on late phase asthmatic response to allergen challenge in asthmatic patients: results of two phase 2a studies. *Lancet* 2007; 370: 1422–1431.
28   McGregor MC, Krings JG, Nair P, *et al.* Role of biologics in asthma. *Am J Respir Crit Care Med* 2019; 199: 433–445.
29   Farzan N, Vijverberg SJ, Kabesch M, *et al.* The use of pharmacogenomics, epigenomics, and transcriptomics to improve childhood asthma management: where do we stand? *Pediatr Pulmonol* 2018; 53: 836–845.
30   Himes BE, Jiang X, Hu R, *et al.* Genome-wide association analysis in asthma subjects identifies SPATS2L as a novel bronchodilator response gene. *PLoS Genet* 2012; 8: e1002824.
31   Farzan N, Vijverberg SJ, Arets HG, *et al.* Pharmacogenomics of inhaled corticosteroids and leukotriene modifiers: a systematic review. *Clin Exp Allergy* 2017; 47: 271–293.
32   Abraham G, Havulinna AS, Bhalala OG, *et al.* Genomic prediction of coronary heart disease. *Eur Heart J* 2016; 37: 3267–3278.
33   Ober C. Asthma genetics in the post-GWAS era. *Ann Am Thorac Soc* 2016; 13: Suppl. 1, S85–S90.

34   Belsky DW, Sears MR, Hancox RJ, *et al.* Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. *Lancet Respir Med* 2013; 1: 453–461.

35   Arabkhazaeli A, Ahmadizar F, Leusink M, *et al.* The association between a genetic risk score for allergy and the risk of developing allergies in childhood – results of the WHISTLER cohort. *Pediatr Allergy Immunol* 2018; 29: 72–77.

36   Hinds DA, McMahon G, Kiefer AK, *et al.* A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* 2013; 45: 907–911.

37   Lehto K, Pedersen NL, Almqvist C, *et al.* Asthma and affective traits in adults: a genetically informative study. *Eur Respir J* 2019; 53: 1802142.

38   Fernandez TD, Mayorga C, Guéant JL, *et al.* Contributions of pharmacogenetics and transcriptomics to the understanding of the hypersensitivity drug reactions. *Allergy* 2014; 69: 150–158.

39   Ewald DA, Malajian D, Krueger JG, *et al.* Meta-analysis derived atopic dermatitis (MADAD) transcriptome defines a robust AD signature highlighting the involvement of atherosclerosis and lipid metabolism pathways. *BMC Med Genomics* 2015; 8: 60.

40   Ghosh D, Ding L, Sivaprasad U, *et al.* Multiple transcriptome data analysis reveals biologically relevant atopic dermatitis signature genes and pathways. *PLoS One* 2015; 10: e0144316.

41   Himes BE, Koziol-White C, Johnson M, *et al.* Vitamin D modulates expression of the airway smooth muscle transcriptome in fatal asthma. *PLoS One* 2015; 10: e0134057.

42   Suárez-Fariñas M, Ungar B, Correa da Rosa J, *et al.* RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications. *J Allergy Clin Immunol* 2015; 135: 1218–1227.

43   Baines KJ, Simpson JL, Wood LG, *et al.* Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *J Allergy Clin Immunol* 2014; 133: 997–1007.

44   Woodruff PG, Modrek B, Choy DF, *et al.* T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180: 388–395.

45   Modena BD, Tedrow JR, Milosevic J, *et al.* Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am J Respir Crit Care Med* 2014; 190: 1363–1372.

46   Modena BD, Bleecker ER, Busse WW, *et al.* Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am J Respir Crit Care Med* 2017; 195: 1449–1463.

47   Östling J, van Geest M, Schofield JPR, *et al.* IL-17-high asthma with features of a psoriasis immunophenotype. *J Allergy Clin Immunol* 2019; 144: 1198–1213.

48   Leaker BR, Malkov VA, Mogg R, *et al.* The nasal mucosal late allergic reaction to grass pollen involves type 2 inflammation (IL-5 and IL-13), the inflammasome (IL-1β), and complement. *Mucosal Immunol* 2017; 10: 408–420.

49   Choy DF, Hart KM, Borthwick LA, *et al.* TH2 and TH17 inflammatory pathways are reciprocally regulated in asthma. *Sci Transl Med* 2015; 7: 301ra129.

50   Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 2015; 16: 197–212.

51   Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018; 50: 96.

52   Seah YFS, Hu H, Merten CA. Microfluidic single-cell technology in immunology and antibody screening. *Mol Aspects Med* 2018; 59: 47–61.

53   Kunz DJ, Gomes T, James KR. Immune cell dynamics unfolded by single-cell technologies. *Front Immunol* 2018; 9: 1435.

54   Neu KE, Tang Q, Wilson PC, *et al.* Single-cell genomics: approaches and utility in immunology. *Trends Immunol* 2017; 38: 140–149.

55   Croote D, Darmanis S, Nadeau KC, *et al.* High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science* 2018; 362: 1306–1309.

56   Chiang D, Chen X, Jones SM, *et al.* Single-cell profiling of peanut-responsive T cells in patients with peanut allergy reveals heterogeneous effector T_H2 subsets. *J Allergy Clin Immunol* 2018; 141: 2107–2120.

57   Saelens W, Cannoodt R, Todorov H, *et al.* A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019; 37: 547–554.

58   Bégin P, Nadeau KC. Epigenetic regulation of asthma and allergic disease. *Allergy Asthma Clin Immunol* 2014; 10: 27.

59   DeVries A, Vercelli D. Epigenetics in allergic diseases. *Curr Opin Pediatr* 2015; 27: 719–723.

60   Harb H, Renz H. Update on epigenetics in allergic disease. *J Allergy Clin Immunol* 2015; 135: 15–24.

61   Brand S, Kesper DA, Teich R, *et al.* DNA methylation of TH1/TH2 cytokine genes affects sensitization and progress of experimental asthma. *J Allergy Clin Immunol* 2012; 129: 1602–1610. e6.

62   Runyon RS, Cachola LM, Rajeshuni N, *et al.* Asthma discordance in twins is linked to epigenetic modifications of T cells. *PLoS One* 2012; 7: e48796.

63   Stefanowicz D, Hackett TL, Garmaroudi FS, *et al.* DNA methylation profiles of airway epithelial cells and PBMCs from healthy, atopic and asthmatic children. *PLoS One* 2012; 7: e44213.

64   Luo Y, Zhou B, Zhao M, *et al.* Promoter demethylation contributes to TSLP overexpression in skin lesions of patients with atopic dermatitis. *Clin Exp Dermatol* 2014; 39: 48–53.

65   Soto-Ramírez N, Arshad SH, Holloway JW, *et al.* The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years. *Clin Epigenetics* 2013; 5: 1.

66   Morales E, Bustamante M, Vilahur N, *et al.* DNA hypomethylation at ALOX12 is associated with persistent wheezing in childhood. *Am J Respir Crit Care Med* 2012; 185: 937–943.

67   Liang L, Willis-Owen SA, Laprise C, *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* 2015; 520: 670–674.

68   Ito K, Caramori G, Lim S, *et al.* Expression and activity of histone deacetylases in human asthmatic airways. *Am J Respir Crit Care Med* 2002; 166: 392–396.

69   Cosío BG, Mann B, Ito K, *et al.* Histone acetylase and deacetylase activity in alveolar macrophages and blood monocytes in asthma. *Am J Respir Crit Care Med* 2004; 170: 141–147.

70   Grausenburger R, Bilic I, Boucheron N, *et al.* Conditional deletion of histone deacetylase 1 in T cells leads to enhanced airway inflammation and increased Th2 cytokine production. *J Immunol* 2010; 185: 3489–3497.

71   Akimova T, Ge G, Golovina T, *et al.* Histone/protein deacetylase inhibitors increase suppressive functions of human FOXP3+ Tregs. *Clin Immunol* 2010; 136: 348–363.

72   Okoye IS, Czieso S, Ktistaki E, *et al.* Transcriptomics identified a critical role for Th2 cell-intrinsic miR-155 in mediating allergy and antihelminth immunity. *Proc Natl Acad Sci USA* 2014; 111: E3081–E3090.

73   Salam MT. Asthma epigenetics. *Adv Exp Med Biol* 2014; 795: 183–199.

74   Moheimani F, Hsu AC, Reid AT, *et al.* The genetic and epigenetic landscapes of the epithelium in asthma. *Respir Res* 2016; 17: 119.

75   von Mutius E. The microbial environment and its influence on asthma prevention in early life. *J Allergy Clin Immunol* 2016; 137: 680–689.

76   Lockett GA, Huoman J, Holloway JW. Does allergy begin in utero? *Pediatr Allergy Immunol* 2015; 26: 394–402.

77   Nicodemus-Johnson J, Myers RA, Sakabe NJ, *et al.* DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* 2016; 1: e90151.

78   Morin A, Laviolette M, Pastinen T, *et al.* Combining omics data to identify genes associated with allergic rhinitis. *Clin Epigenetics* 2017; 9: 3.

79   Xu CJ, Bonder MJ, Söderhäll C, *et al.* The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics* 2017; 18: 25.

80   Lynch JP, Sikder MA, Curren BF, *et al.* The influence of the microbiome on early-life severe viral lower respiratory infections and asthma – food for thought? *Front Immunol* 2017; 8: 156.

81   Stiemsma LT, Turvey SE. Asthma and the microbiome: defining the critical window in early life. *Allergy Asthma Clin Immunol* 2017; 13: 3.

82   Fu L, Wang C, Wang Y. Seafood allergen-induced hypersensitivity at the microbiota-mucosal site: implications for prospective probiotic use in allergic response regulation. *Crit Rev Food Sci Nutr* 2018; 58: 1512–1525.

83   West CE, Renz H, Jenmalm MC, *et al.* The gut microbiota and inflammatory noncommunicable diseases: associations and potentials for gut microbiota therapies. *J Allergy Clin Immunol* 2015; 135: 3–13.

84   Huang YJ, Boushey HA. The microbiome in asthma. *J Allergy Clin Immunol* 2015; 135: 25–30.

85   Teo SM, Mok D, Pham K, *et al.* The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 2015; 17: 704–715.

86   Sirisinha S. The potential impact of gut microbiota on your health: current status and future challenges. *Asian Pac J Allergy Immunol* 2016; 34: 249–264.

87   Chotirmall SH, Gellatly SL, Budden KF, *et al.* Microbiomes in respiratory health and disease: an Asia-Pacific perspective. *Respirology* 2017; 22: 240–250.

88   Kang YB, Cai Y, Zhang H. Gut microbiota and allergy/asthma: from pathogenesis to new therapeutic strategies. *Allergol Immunopathol* 2017; 45: 305–309.

89   Yang X, Jiang Y, Wang C. Does IL-17 respond to the disordered lung microbiome and contribute to the neutrophilic phenotype in asthma? *Mediators Inflamm* 2016; 2016: 6470364.

90   Azad MB, Konya T, Guttman DS, *et al.* Infant gut microbiota and food sensitization: associations in the first year of life. *Clin Exp Allergy* 2015; 45: 632–643.

91   Blázquez AB, Berin MC. Microbiome and food allergy. *Transl Res* 2017; 179: 199–203.

92   Holt PG, Sly PD. Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. *Nat Med* 2012; 18: 726–735.

93   Calışkan M, Bochkov YA, Kreiner-Møller E, *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* 2013; 368: 1398–1407.

94   Bochkov YA, Watters K, Ashraf S, *et al.* Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc Natl Acad Sci USA* 2015; 112: 5485–5490.

95   Basnet S, Bochkov YA, Brockman-Schneider RA, *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.

96   Gern JE, Pappas T, Visness CM, *et al.* Comparison of the etiology of viral respiratory illnesses in inner-city and suburban infants. *J Infect Dis* 2012; 206: 1342–1349.

97   Man WH, de Steenhuijsen Piters WA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 2017; 15: 259–270.

98   Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for environmental health research. *Eur Respir Rev* 2016; 25: 124–129.

99   North ML, Brook JR, Lee EY, *et al.* The Kingston Allergy Birth Cohort: exploring parentally reported respiratory outcomes through the lens of the exposome. *Ann Allergy Asthma Immunol* 2017; 118: 465–473.

100   Miyata J, Arita M. Role of omega-3 fatty acids and their metabolites in asthma and allergic diseases. *Allergol Int* 2015; 64: 27–34.

101   Magnusson J, Ekström S, Kull I, *et al.* Polyunsaturated fatty acids in plasma at 8 years and subsequent allergic disease. *J Allergy Clin Immunol* 2018; 142: 510–516.

102   Capozzi F, Bordoni A. Foodomics: a new comprehensive approach to food and nutrition. *Genes Nutr* 2013; 8: 1–4.

103   Lynch SV, Wood RA, Boushey H, *et al.* Effects of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban children. *J Allergy Clin Immunol* 2014; 134: 593–601.

104   Marchetti P, Pesce G, Villani S, *et al.* Pollen concentrations and prevalence of asthma and allergic rhinitis in Italy: evidence from the GEIRD study. *Sci Total Environ* 2017; 584–585: 1093–1099.

105   Du Toit G, Roberts G, Sayre PH, *et al.* Randomized trial of peanut consumption in infants at risk for peanut allergy. *N Engl J Med* 2015; 372: 803–813.

106   Benedé S, Blázquez AB, Chiang D, *et al.* The rise of food allergy: environmental factors and emerging treatments. *EBioMedicine* 2016; 7: 27–34.

107   Shankardass K, Jerrett M, Dell SD, *et al.* Spatial analysis of exposure to traffic-related air pollution at birth and childhood atopic asthma in Toronto, Ontario. *Health Place* 2015; 34: 287–295.

108   Vincent MJ, Bernstein JA, Basketter D, *et al.* Chemical-induced asthma and the role of clinical, toxicological, exposure and epidemiological research in regulatory and hazard characterization approaches. *Regul Toxicol Pharmacol* 2017; 90: 126–132.

109   Lombardo LJ, Balmes JR. Occupational asthma: a review. *Environ Health Perspect* 2000; 108: Suppl. 4, 697–704.

110   Kimber I, Dearman RJ, Basketter DA, *et al.* Chemical respiratory allergy: reverse engineering an adverse outcome pathway. *Toxicology* 2014; 318: 32–39.

111   Clark A, Mach N. Role of vitamin D in the hygiene hypothesis: the interplay between vitamin D, vitamin D receptors, gut microbiota, and immune response. *Front Immunol* 2016; 7: 627.

112   Wjst M. Linking vitamin D, the microbiome and allergy. *Allergy* 2017; 72: 329–330.

113   Rossi R, De Palma A, Benazzi L, *et al.* Biomarker discovery in asthma and COPD by proteomic approaches. *Proteomics Clin Appl* 2014; 8: 901–915.

114   Terracciano R, Pelaia G, Preianò M, *et al.* Asthma and COPD proteomics: current approaches and future directions. *Proteomics Clin Appl* 2015; 9: 203–220.

115   Teran LM, Montes-Vizuet R, Li X, *et al.* Respiratory proteomics: from descriptive studies to personalized medicine. *J Proteome Res* 2015; 14: 38–50.

116   Tomazic PV, Birner-Gruenberger R, Leitner A, *et al.* Seasonal proteome changes of nasal mucus reflect perennial inflammatory response and reduced defence mechanisms and plasticity in allergic rhinitis. *J Proteomics* 2016; 133: 153–160.

117   Schofield JPR, Burg D, Nicholas B, *et al.* Stratification of asthma phenotypes by airway proteomic signatures. *J Allergy Clin Immunol* 2019; 144: 70–82.

118   Ferreira MA. Cytokine expression in allergic inflammation: systematic review of *in vivo* challenge studies. *Mediators Inflamm* 2003; 12: 259–267.

119   Di Girolamo F, Muraca M, Mazzina O, *et al.* Proteomic applications in food allergy: food allergenomics. *Curr Opin Allergy Clin Immunol* 2015; 15: 259–266.

120   Chan TF, Ji KM, Yim AK, *et al.* The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015; 135: 539–548.

121   Choopong J, Reamtong O, Sookrung N, *et al.* Proteome, allergenome, and novel allergens of house dust mite, *Dermatophagoides farinae*. *J Proteome Res* 2016; 15: 422–430.

122   Oseroff C, Christensen LH, Westernberg L, *et al.* Immunoproteomic analysis of house dust mite antigens reveals distinct classes of dominant T cell antigens according to function and serological reactivity. *Clin Exp Allergy* 2017; 47: 577–592.

123   Campbell BC, Gilding EK, Timbrell V, *et al.* Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *J Allergy Clin Immunol* 2015; 135: 133–142.

124   Ghosh N, Sircar G, Saha B, *et al.* Search for allergens from the pollen proteome of sunflower (*Helianthus annuus* L.): a major sensitizer for respiratory allergy patients. *PLoS One* 2015; 10: e0138992.

125   Tiotiu A, Brazdova A, Longé C, *et al. Urtica dioica* pollen allergy: clinical, biological, and allergomics analysis. *Ann Allergy Asthma Immunol* 2016; 117: 527–534.

126   Scrivo R, Casadei L, Valerio M, *et al.* Metabolomics approach in allergic and rheumatic diseases. *Curr Allergy Asthma Rep* 2014; 14: 445.

127   Kelly RS, Dahlin A, McGeachie MJ, *et al.* Asthma metabolomics and the potential for integrative omics in research and the clinic. *Chest* 2017; 151: 262–277.

128   Villaseñor A, Rosace D, Obeso D, *et al.* Allergic asthma: an overview of metabolomic strategies leading to the identification of biomarkers in the field. *Clin Exp Allergy* 2017; 47: 442–456.

129   Kasuga K, Suga T, Mano N. Bioanalytical insights into mediator lipidomics. *J Pharm Biomed Anal* 2015; 113: 151–162.

130   Kunisawa J, Kiyono H. Sphingolipids and epoxidized lipid metabolites in the control of gut immunosurveillance and allergy. *Front Nutr* 2016; 3: 3.

131   Pendergrass SA, Brown-Gentry K, Dudek S, *et al.* Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet* 2013; 9: e1003087.

132   Karaca S, Civelek E, Karaca M, *et al.* Allergy-specific phenome-wide association study for immunogenes in Turkish children. *Sci Rep* 2016; 6: 33152.

133   Vigh-Conrad KA, Conrad DF, Preuss D. A protein allergen microarray detects specific IgE to pollen surface, cytoplasmic, and commercial allergen extracts. *PLoS One* 2010; 5: e10174.

134   Mari A, Scala E, Alessandri C. The IgE-microarray testing in atopic dermatitis: a suitable modern tool for the immunological and clinical phenotyping of the disease. *Curr Opin Allergy Clin Immunol* 2011; 11: 438–444.

135   Pomponi D, Bernardi ML, Liso M, *et al.* Allergen micro-bead array for IgE detection: a feasibility study using allergenic molecules tested on a flexible multiplex flow cytometric immunoassay. *PLoS One* 2012; 7: e35697.

136   Hosoki K, Ying S, Corrigan C, *et al.* Analysis of a panel of 48 cytokines in BAL fluids specifically identifies IL-8 levels as the only cytokine that distinguishes controlled asthma from uncontrolled asthma, and correlates inversely with FEV$_1$. *PLoS One* 2015; 10: e0126035.

137   Blanchard C, Stucke EM, Rodriguez-Jimenez B, *et al.* A striking local esophageal cytokine expression profile in eosinophilic esophagitis. *J Allergy Clin Immunol* 2011; 127: 208–217.

138   Kidd BA, Peters LA, Schadt EE, *et al.* Unifying immunology with informatics and multiscale biology. *Nat Immunol* 2014; 15: 118–127.

139   Biancotto A, McCoy JP. Studying the human immunome: the complexity of comprehensive leukocyte immunophenotyping. *Curr Top Microbiol Immunol* 2014; 377: 23–60.

140   Olnes MJ, Kotliarov Y, Biancotto A, *et al.* Effects of systemically administered hydrocortisone on the human immunome. *Sci Rep* 2016; 6: 23002.

141   Yao Y, Welp T, Liu Q, *et al.* Multiparameter single cell profiling of airway inflammatory cells. *Cytometry B Clin Cytom* 2017; 92: 12–20.

142   Olin A, Henckel E, Chen Y, *et al.* Stereotypic immune system development in newborn children. *Cell* 2018; 174: 1277–1292.

143   Lee AH, Shannon CP, Amenyogbe N, *et al.* Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nat Commun* 2019; 10: 1092.

144   Alpert A, Pickman Y, Leipold M, *et al.* A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat Med* 2019; 25: 487–495.

145   Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol* 2016; 16: 626–638.

146   LoMauro A, Aliverti A. Sex differences in respiratory function. *Breathe* 2018; 14: 131–140.

147   Yung JA, Fuseini H, Newcomb DC. Hormones, sex, and asthma. *Ann Allergy Asthma Immunol* 2018; 120: 488–494.

148   Minelli C, van der Plaat DA, Leynaert B, *et al.* Age at puberty and risk of asthma: a Mendelian randomisation study. *PLoS Med* 2018; 15: e1002634.

149   Porcu E, Rüeger S, Lepik K, *et al.* Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun* 2019; 10: 3300.

150   Gref A, Merid SK, Gruzieva O, *et al.* Genome-wide interaction analysis of air pollution exposure and childhood asthma with functional follow-up. *Am J Respir Crit Care Med* 2017; 195: 1373–1383.

151   Vonk JM, Scholtens S, Postma DS, *et al.* Adult onset asthma and interaction between genes and active tobacco smoking: the GABRIEL consortium. *PLoS One* 2017; 12: e0172716.

152   Sordillo JE, Kelly R, Bunyavanich S, *et al.* Genome-wide expression profiles identify potential targets for gene-environment interactions in asthma severity. *J Allergy Clin Immunol* 2015; 136: 885–892.

153   Li L, Kabesch M, Bouzigon E, *et al.* Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. *Front Genet* 2013; 4: 103.

154   Dai X, Bowatte G, Lowe AJ, *et al.* Do glutathione S-transferase genes modify the link between indoor air pollution and asthma, allergies, and lung function? A systematic review. *Curr Allergy Asthma Rep* 2018; 18: 20.

155   Kabesch M, Hoefler C, Carr D, *et al.* Glutathione S transferase deficiency and passive smoking increase childhood asthma. *Thorax* 2004; 59: 569–573.

156   Polonikov AV, Ivanov VP, Solodilova MA. Genetic variation of genes for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the importance of gene-gene and gene-environment interactions for disease susceptibility. *J Hum Genet* 2009; 54: 440–449.

157   Hur GY, Choi SJ, Shin SY, *et al.* Update on the pathogenic mechanisms of isocyanate-induced asthma. *World Allergy Organ J* 2008; 1: 15–18.

158   Lee YL, Lin YC, Lee YC, *et al.* Glutathione S-transferase P1 gene polymorphism and air pollution as interactive risk factors for childhood asthma. *Clin Exp Allergy* 2004; 34: 1707–1713.

159   Wagener AH, Zwinderman AH, Luiten S, *et al.* dsRNA-induced changes in gene expression profiles of primary nasal and bronchial epithelial cells from patients with asthma, rhinitis and controls. *Respir Res* 2014; 15: 9.

160   Wesolowska-Andersen A, Everman JL, Davidson R, *et al.* Dual RNA-seq reveals viral infections in asthmatic children without respiratory illness which are associated with changes in the airway transcriptome. *Genome Biol* 2017; 18: 12.

161   Zhao F, Durner J, Winkler JB, *et al.* Pollen of common ragweed (*Ambrosia artemisiifolia* L.): Illumina-based *de novo* sequencing and differential transcript expression upon elevated NO_2/O_3. *Environ Pollut* 2017; 224: 503–514.

162   Jahreis S, Trump S, Bauer M, *et al.* Maternal phthalate exposure promotes allergic airway inflammation over 2 generations through epigenetic modifications. *J Allergy Clin Immunol* 2018; 141: 741–753.

163   Nath AP, Ritchie SC, Byars SG, *et al.* An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. *Genome Biol* 2017; 18: 146.

164   Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29.

165   Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000; 28: 235–242.

166   Apweiler R, Bairoch A, Wu CH, *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 2004; 32: D115–D119.

167   Inouye M, Kettunen J, Soininen P, *et al.* Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* 2010; 6: 441.

168   Inouye M, Silander K, Hamalainen E, *et al.* An immune response network associated with blood lipid levels. *PLoS Genet* 2010; 6: e1001113.

169   Sajda P. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng* 2006; 8: 537–565.

170   Tarca AL, Carey VJ, Chen XW, *et al.* Machine learning and its applications to biology. *PLoS Comput Biol* 2007; 3: e116.

171   Sommer C, Gerlich DW. Machine learning in cell biology – teaching computers to recognize phenotypes. *J Cell Sci* 2013; 126: 5529–5539.

172   Tan P-N, Kumar V, Steinbach M. Introduction to Data Mining. Boston, Pearson Addison Wesley, 2005.

173   Clark NR, Ma'ayan A. Introduction to statistical methods to analyze large data sets: principal components analysis. *Sci Signal* 2011; 4: tr3.

174   Everitt B. Cluster Analysis. 3rd edn. London, Arnold, 1993.

175   Wenzel S. Severe asthma: from characteristics to phenotypes to endotypes. *Clin Exp Allergy* 2012; 42: 650–658.

176   Hastie AT, Moore WC, Meyers DA, *et al.* Analyses of asthma severity phenotypes and inflammatory proteins in subjects stratified by sputum granulocytes. *J Allergy Clin Immunol* 2010; 125: 1028–1036. e13.

177   Raedler D, Ballenberger N, Klucker E, *et al.* Identification of novel immune phenotypes for allergic and nonallergic childhood asthma. *J Allergy Clin Immunol* 2015; 135: 81–91.

178   Spycher BD, Silverman M, Kuehni CE. Phenotypes of childhood asthma: are they real? *Clin Exp Allergy* 2010; 40: 1130–1141.

179   Bisgaard H, Bønnelykke K. Long-term studies of the natural history of asthma in childhood. *J Allergy Clin Immunol* 2010; 126: 187–197.

180   Lodge CJ, Zaloumis S, Lowe AJ, *et al.* Early-life risk factors for childhood wheeze phenotypes in a high-risk birth cohort. *J Pediatr* 2014; 164: 289–294.

181   Moore WC, Meyers DA, Wenzel SE, *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315–323.

182   Moore WC, Hastie AT, Li X, *et al.* Sputum neutrophil counts are associated with more severe asthma phenotypes using cluster analysis. *J Allergy Clin Immunol* 2014; 133: 1557–1563.

183   Chung KF, Adcock IM. Clinical phenotypes of asthma should link up with disease mechanisms. *Curr Opin Allergy Clin Immunol* 2015; 15: 56–62.

184   Kuo CS, Pavlidis S, Loza M, *et al.* T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J* 2017; 49: 1602135.

185   Weatherall M, Travers J, Shirtcliffe PM, *et al.* Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J* 2009; 34: 812–818.

186   Simpson A, Tan VY, Winn J, *et al.* Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med* 2010; 181: 1200–1206.

187   Bousquet J, Anto JM, Akdis M, *et al.* Paving the way of systems biology and precision medicine in allergic diseases: the MeDALL success story: Mechanisms of the Development of ALLergy; EU FP7-CP-IP; Project No: 261357; 2010–2015. *Allergy* 2016; 71: 1513–1525.

188   Pavlidis S, Sun K, Guo Y, *et al.* Longitudinal follow-up of sputum transcriptomic-associated clusters in asthma from the U-BIOPRED cohort. *Am J Respir Crit Care Med* 2018; 197: A5957.

189   Boudier A, Curjuric I, Basagaña X, *et al.* Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. *Am J Respir Crit Care Med* 2013; 188: 550–560.

190   Belgrave DC, Granell R, Simpson A, *et al.* Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med* 2014; 11: e1001748.

191   Garden FL, Simpson JM, Mellis CM, *et al.* Change in the manifestations of asthma and asthma-related traits in childhood: a latent transition analysis. *Eur Respir J* 2016; 47: 499–509.

192   Tang HH, Teo SM, Belgrave DC, *et al.* Trajectories of childhood immune development and respiratory health relevant to asthma and allergy. *Elife* 2018; 7: e35856.

193   Tang HHF, Lang A, Evans M, *et al.* The early-life nasopharyngeal microbiome interacts with allergic and non-allergic mechanisms of childhood wheeze. ASCIA 2019 Conference Abstracts: Poster 1. *Intern Med J* 2019; 49: Suppl. 4, 22.

194   Himes BE, Hunninghake GM, Baurley JW, *et al.* Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* 2009; 84: 581–593.

195   Hinks TS, Zhou X, Staples KJ, *et al.* Innate and adaptive T cells in asthmatic patients: relationship to severity and disease mechanisms. *J Allergy Clin Immunol* 2015; 136: 323–333.

196   Chu JH, Weiss ST, Carey VJ, *et al.* A graphical model approach for inferring large-scale networks integrating gene expression and genetic polymorphism. *BMC Syst Biol* 2009; 3: 55.

197   Hao K, Bossé Y, Nickle DC, *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8: e1003029.

198   Troy NM, Hollams EM, Holt PG, *et al.* Differential gene network analysis for the identification of asthma-associated therapeutic targets in allergen-specific T-helper memory responses. *BMC Med Genomics* 2016; 9: 9.

199   Kim YW, Singh A, Shannon CP, *et al.* Transcriptional networks in whole blood of asthmatics. *Allergy Asthma Clin Immunol* 2014; 10: A58.

200   Pillai RR, Divekar R, Brasier A, *et al.* Strategies for molecular classification of asthma using bipartite network analysis of cytokine expression. *Curr Allergy Asthma Rep* 2012; 12: 388–395.

201   Perkins JR, Barrionuevo E, Ranea JA, *et al.* Systems biology approaches to enhance our understanding of drug hypersensitivity reactions. *Clin Exp Allergy* 2014; 44: 1461–1472.

202   Dahlin A, Tantisira KG. Integrative systems biology approaches in asthma pharmacogenomics. *Pharmacogenomics* 2012; 13: 1387–1404.

203   Lynn DJ, Winsor GL, Chan C, *et al.* InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* 2008; 4: 218.

204   Sircar G, Saha B, Bhattacharya SG, *et al.* Allergic asthma biomarkers using systems approaches. *Front Genet* 2014; 4: 308.

205   Zissler UM, Esser-von Bieren J, Jakwerth CA, *et al.* Current and future biomarkers in allergic asthma. *Allergy* 2016; 71: 475–494.

206   Kips JC, Tournoy KG, Pauwels RA. Gene knockout models of asthma. *Am J Respir Crit Care Med* 2000; 162: S66–S70.

207   Zhang Y, Wu Y, Qi H, *et al.* A new antagonist for CCR4 attenuates allergic lung inflammation in a mouse model of asthma. *Sci Rep* 2017; 7: 15038.

208   Singh D, Cadden P, Hunter M, *et al.* Inhibition of the asthmatic allergen challenge response by the CRTH2 antagonist OC000459. *Eur Respir J* 2013; 41: 46–52.

209   Höfer T, Nathansen H, Löhning M, *et al.* GATA-3 transcriptional imprinting in Th2 lymphocytes: a mathematical model. *Proc Natl Acad Sci USA* 2002; 99: 9364–9368.

210   Lauzon AM, Bates JH, Donovan G, *et al.* A multi-scale approach to airway hyperresponsiveness: from molecule to organ. *Front Physiol* 2012; 3: 191.

211   Narang V, Decraene J, Wong SY, *et al.* Systems immunology: a survey of modeling formalisms, applications and simulation tools. *Immunol Res* 2012; 53: 251–265.

212   Auffray C, Adcock IM, Chung KF, *et al.* An integrative systems biology approach to understanding pulmonary diseases. *Chest* 2010; 137: 1410–1416.

213   Regev A, Teichmann S, Lander ES, *et al.* The human cell atlas. *Elife* 2017; 6: e27041.

214   Vercelli D, Martinez FD. The Faustian bargain of genetic association studies: bigger might not be better, or at least it might not be good enough. *J Allergy Clin Immunol* 2006; 117: 1303–1305.

215   Huang Qin Q, Ritchie SC, Brozynska M, *et al.* Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res* 2018; 46: e133.

216   Belgrave D, Henderson J, Simpson A, *et al.* Disaggregating asthma: big investigation *versus* big data. *J Allergy Clin Immunol* 2017; 139: 400–407.

217   Chen JH, Asch SM. Machine learning and prediction in medicine – beyond the peak of inflated expectations. *N Engl J Med* 2017; 376: 2507–2509.

218   Kupczyk M, Dahlén B, Sterk PJ, *et al.* Stability of phenotypes defined by physiological variables and biomarkers in adults with asthma. *Allergy* 2014; 69: 1198–1204.

219   Deliu M, Sperrin M, Belgrave D, *et al.* Identification of asthma subtypes using clustering methodologies. *Pulm Ther* 2016; 2: 19–41.

220   Lazic N, Roberts G, Custovic A, *et al.* Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. *Allergy* 2013; 68: 764–770.

221   Bromham L, Dinnage R, Hua X. Interdisciplinary research has consistently lower funding success. *Nature* 2016; 534: 684–687.

222   Sparks R, Lau WW, Tsang JS. Expanding the immunology toolbox: embracing public-data reuse and crowdsourcing. *Immunity* 2016; 45: 1191–1204.

223   Sittka A, Vera J, Lai X, *et al.* Asthma phenotyping, therapy, and prevention: what can we learn from systems biology? *Pediatr Res* 2013; 73: 543–552.

224   Galli SJ. Toward precision medicine and health: opportunities and challenges in allergic diseases. *J Allergy Clin Immunol* 2016; 137: 1289–1300.

225   Bigler J, Boedigheimer M, Schofield JP, *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.