

RESEARCH

Open Access



# TWIST1 DNA methylation is a cell marker of airway and parenchymal lung fibroblasts that are differentially methylated in asthma

Rachel L. Clifford<sup>1,3\*</sup> , Chen Xi Yang<sup>2</sup>, Nick Fishbane<sup>2</sup>, Jamie Patel<sup>1</sup>, Julia L. Maclsaac<sup>3</sup>, Lisa M. McEwen<sup>3</sup>, Sean T. May<sup>4</sup>, Marcos Castellanos-Uribe<sup>4</sup>, Parameswaran Nair<sup>5</sup>, Ma'en Obeidat<sup>2,6</sup>, Michael S. Kobor<sup>3</sup>, Alan J. Knox<sup>1</sup> and Tillie-Louise Hackett<sup>2,6</sup>

## Abstract

**Background:** Mesenchymal fibroblasts are ubiquitous cells that maintain the extracellular matrix of organs. Within the lung, airway and parenchymal fibroblasts are crucial for lung development and are altered with disease, but it has been difficult to understand their roles due to the lack of distinct molecular markers. We studied genome-wide DNA methylation and gene expression in airway and parenchymal lung fibroblasts from healthy and asthmatic donors, to identify a robust cell marker and to determine if these cells are molecularly distinct in asthma.

**Results:** Airway ( $N = 8$ ) and parenchymal ( $N = 15$ ) lung fibroblasts from healthy individuals differed in the expression of 158 genes, and DNA methylation of 3936 CpGs (Bonferroni adjusted  $p$  value  $< 0.05$ ). Differential DNA methylation between cell types was associated with differential expression of 42 genes, but no single DNA methylation CpG feature (location, effect size, number) defined the interaction. Replication of gene expression and DNA methylation in a second cohort identified TWIST1 gene expression, DNA methylation and protein expression as a cell marker of airway and parenchymal lung fibroblasts, with DNA methylation having 100% predictive discriminatory power. DNA methylation was differentially altered in parenchymal (112 regions) and airway fibroblasts (17 regions) with asthmatic status, with no overlap between regions.

**Conclusions:** Differential methylation of TWIST1 is a robust cell marker of airway and parenchymal lung fibroblasts. Airway and parenchymal fibroblast DNA methylation are differentially altered in individuals with asthma, and the role of both cell types should be considered in the pathogenesis of asthma.

**Keywords:** DNA methylation, Gene expression, Cell marker, Lung, Fibroblast, Airway, Parenchyma, Asthma

\* Correspondence: [r.clifford@nottingham.ac.uk](mailto:r.clifford@nottingham.ac.uk)

<sup>1</sup>Nottingham NIHR Biomedical Research Centre, Nottingham MRC Molecular Pathology Node, Division of Respiratory Medicine, University of Nottingham, Nottingham University Hospitals NHS Trust, City Hospital, Nottingham, UK

<sup>3</sup>BC Children's Hospital Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Fibroblasts are mesenchymal cells found in the stroma of many tissues and organs throughout the body that are essential for the secretion and maintenance of the extracellular matrix (ECM), in the settings of tissue homeostasis, repair and disease. Fibroblasts are traditionally defined by their spindle-shaped morphology and the relatively non-specific expression of vimentin and S100a4 (Fibroblast specific protein (FSP)-1). However, a lack of lineage markers has hindered the understanding of the specific origins and functions of different fibroblast populations [1]. Gene expression profiling of fibroblasts taken from 35 different anatomic locations, including the skin, lung, aorta, liver, skeletal muscle and prostate, has shown fibroblasts from different organs are molecularly distinct from each other [1, 2]. However, the segregation of fibroblasts by anatomical organ relied on panels of large numbers of genes [2] with few genes distinct to fibroblasts from different organs [1, 2]. Thus, a single gene expression marker of fibroblast anatomical location has been elusive. The problem is more complex in organs such as the lung that contain discrete tissues such as the conducting airways and parenchymal tissue, with distinct fibroblast populations in each structure. Airway and parenchymal fibroblasts have been shown to be morphologically [3, 4] and molecularly distinct [5]. Such variation in fibroblast biology is important to understand, as many common lung diseases such as asthma present with fibrosis in airway versus parenchymal tissues [6, 7], highlighting a need to understand the role of these distinct fibroblast populations.

DNA methylation is commonly studied in the context of disease dysfunction; however, the understanding of its

role in regulating normal tissue-specific genome function has only gained momentum in recent years. A vast amount of DNA methylation data now exists across a wide variety of specific cell types, with several studies identifying differentially methylated regions that are both tissue- and organ-specific [8–11]. Further, given the stable nature of DNA methylation relative to gene expression, it may provide better molecular markers to distinguish specific tissues and cell types.

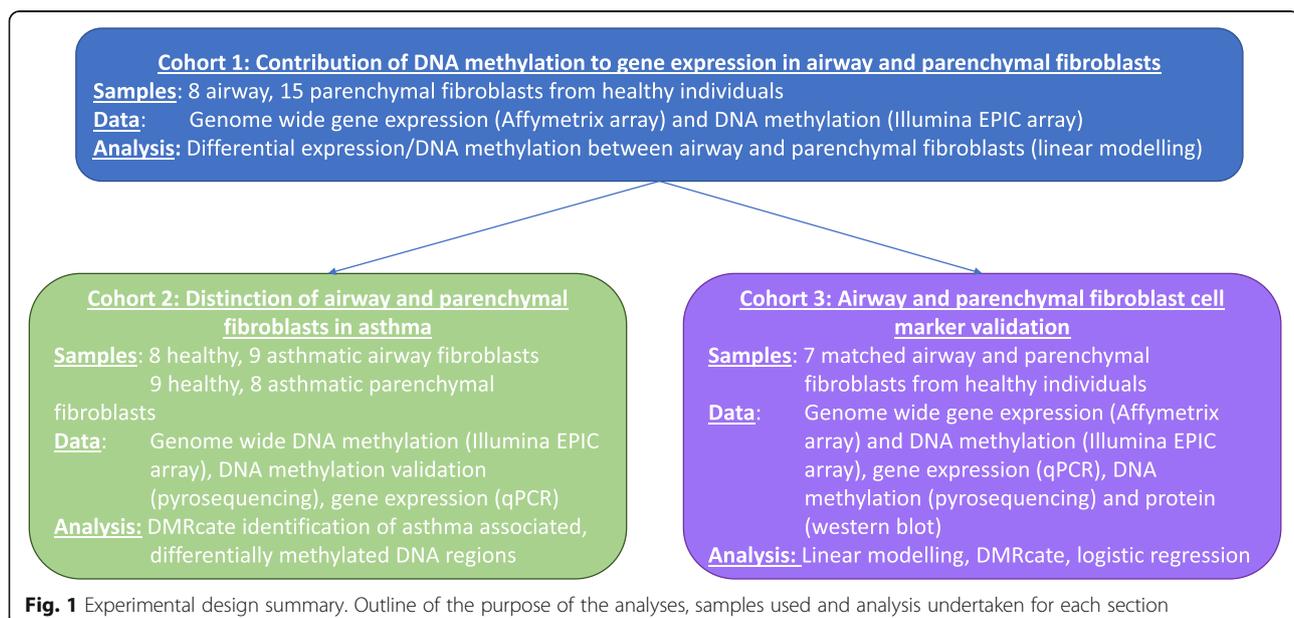
In the present study (Fig. 1), genome-wide DNA methylation and gene expression were profiled in airway and parenchymal fibroblasts from healthy individuals, to understand if DNA methylation contributes to the heterogeneity of lung fibroblasts and can be used as a robust cell marker. Lastly, we assessed whether DNA methylation is altered in airway and parenchymal fibroblasts isolated from individuals with asthma.

## Results

### Airway and parenchymal fibroblasts exhibit different gene expression profiles

We compared genome-wide gene expression between airway fibroblasts ( $n = 8$ ) and parenchymal fibroblasts ( $n = 15$ ) collected from healthy individuals with no respiratory disease or medication history. The demographics of the 23 healthy subjects are provided in Table 1.

We found 3624 probe sets, representing 2619 unique genes, that were significant and differentially expressed between airway and parenchymal lung fibroblasts (Fig. 2a, dark grey points, Benjamini-Hochberg false discovery rate (FDR) < 0.05). Of these, 963 genes had a greater than 1.5-fold difference in gene expression, with 526 genes having higher expression in airway fibroblasts (Fig. 2a, blue





(See figure on previous page.)

**Fig. 2** Gene expression and DNA methylation differs between airway and parenchymal fibroblasts. Airway fibroblasts were used as reference. **a** Plot of 53,617 gene expression probes used in the analysis. Red/Blue = Benjamini-Hochberg FDR < 0.05. The horizontal dotted line indicates the  $p$  value corresponding to Benjamini-Hochberg FDR < 0.05. Red = greater expression in parenchymal fibroblasts, blue = greater expression in airway fibroblasts. **b** Plot of 414,592 CpG probes used in the analysis. Red/Blue = Bonferroni adjusted  $p$  value < 0.05. Airway fibroblasts were used as reference. Red = greater methylation in parenchymal fibroblasts, blue = greater methylation in airway fibroblasts. The horizontal dotted line indicates the  $p$ -value corresponding to adjusted  $p$  value 0.05. **c** Distribution of DNA methylation (beta) difference between airway and parenchymal fibroblasts in significant probes (Bonferroni adjusted  $p$  value < 0.05). Red line = difference in beta of 0.2, which approximates a 20% change, green line = beta difference of 0.5. Bars in red indicate an absolute beta difference greater than 0.5 between airway and lung fibroblasts (differential methylation > 50%). **d** Gene feature type for all analysis probes, all significant probes, significant probes with a difference in methylation greater than 50%. **e** CpG density for all analysis probes, all significant probes, significant probes with a difference in methylation greater than 50%. **f** Gene feature type for all analysis probes, all significant probes, significant probes with a decrease in methylation of greater than 50% and significant probes with an increase in methylation of greater than 50%. **g** CpG density type for all analysis probes, all significant probes, significant probes with a decrease in methylation of greater than 50% and significant probes with an increase in methylation of greater than 50%. UTR, untranslated region; TSS, transcription start site; Island, > 200 bp with > 50% GC percentage; Shore, up to 2 kb from a CpG island; N\_Shore, North shore upstream of CpG island; S\_Shore downstream of CpG island; Shelf, 2–4 kb from a CpG island; OpenSea, regions of the genome without any enrichment of CpG content [12, 13]

points), and 437 genes being more highly expressed in parenchymal fibroblasts (Fig. 2a, red points). Our findings confirm previous observations that normal airway and parenchymal fibroblasts exhibit different gene expression profiles [5]. Of the 963 differentially expressed genes (> 1.5 fold change), only 123 (detailed in Supplemental Table 1) were replicated from the 775 unique genes previously identified as a gene signature of airway and parenchymal fibroblasts [5], demonstrating the transient and variable nature of gene expression.

#### Airway and parenchymal fibroblasts exhibit different DNA methylation profiles

Having confirmed that gene expression profiles differed between airway and parenchymal fibroblasts, we next assessed if DNA methylation could provide further molecular distinction of the two cell types. In a site-by-site analysis, DNA methylation identified 3936 CpGs that were significantly differentially methylated between airway and parenchymal lung fibroblasts (Bonferroni adjusted  $p$  value < 0.05) (Fig. 2b). To understand the characteristics of the CpGs differentially methylated in parenchymal and airway lung fibroblasts, we considered the direction of difference in CpG methylation (Fig. 2b), the magnitude of CpG methylation difference (Fig. 2c), CpG genomic location (Fig. 2d) and CpG density of the target site location (Fig. 2e). The directional effect of DNA methylation was balanced between the two cell types, with 2088 CpG sites (53%) being more methylated (Fig. 2b, red points) and 1848 (47%) CpG sites being less methylated in parenchymal compared to airway fibroblasts (Fig. 2b, blue points). The effect sizes in terms of DNA methylation differences between airway and parenchymal lung fibroblasts were notable, 3234 (82.16%) CpGs had an absolute beta difference of greater than 0.2 (red line, Fig. 2c) and 240 (6.09%) CpGs had an absolute beta difference of greater than 0.5 (50% methylation difference) (green line, Fig. 2c). Differently methylated

CpGs were located across all regions of the genome (Fig. 2d) and within all classifications of CpG density (Fig. 2e); however, they were located primarily in gene body regions (Fig. 2d), and in regions of open sea CpG density (regions of the genome without any enrichment of CpG content [12, 13] (Fig. 2e)). Enrichment within these locations differed significantly from the distribution of the full data set (gene feature,  $\chi^2$   $p$  value < 0.017 (Fig. 2d); CpG density,  $\chi^2$   $p$  value < 0.0188, (Fig. 2e)).

To assess whether CpG genomic positioning had any bearing on the magnitude of DNA methylation difference between airway and parenchymal lung fibroblasts, we assessed enrichment to CpG location for sites with a greater than 50% difference in DNA methylation. Enrichment in gene bodies was maintained for sites with a beta difference of > 0.5 (Fig. 2d,  $\chi^2$   $p$  value < 0.0012); however, enrichment in open sea regions was lost (Fig. 2e,  $\chi^2$   $p$  value = 0.7124). This suggested that the CpG genomic location of differentially methylated CpGs was not related to effect size but rather sites of smaller effect size contributed significantly to open sea enrichment. To assess whether genomic positioning of a CpG had any influence on the direction of DNA methylation difference between airway and parenchymal fibroblasts, we assessed enrichment of CpG location by direction difference. Similarly, gene body enrichment was maintained regardless of whether a higher ( $\chi^2$   $p$  value = 0.0256) or lower ( $\chi^2$   $p$  value = 0.0098) level of DNA methylation was observed in parenchymal fibroblasts relative to airway fibroblasts (Fig. 2f). However, enrichment in open sea regions was observed only for CpGs with a higher level of DNA methylation in parenchymal fibroblasts relative to airway fibroblasts (Fig. 2g; decrease  $\chi^2$   $p$  value = 0.0621, increase  $\chi^2$   $p$  value = 0.0022) suggesting some propensity for open sea DNA methylation in parenchymal fibroblasts relative to airway fibroblasts.

Gene set enrichment testing using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes

(KEGG) and Reactome identified significant enrichment of CpGs in genes involved in extracellular matrix (ECM) organization, constitution and degradation, cell-cell communication, cell adhesion and muscle contraction (Supplemental Table 2), consistent with previous reports of the different fibrotic functions of fibroblast cell types [3, 4].

#### DNA methylation is associated with differential gene expression in airway and parenchymal fibroblasts

To better understand the relationship between differential gene expression and DNA methylation, we integrated the two datasets. To be consistent with the DNA methylation data, we used Bonferroni-corrected gene expression data, which resulted in 178 probe sets, representing 158 unique genes that were differentially expressed between airway and parenchymal lung fibroblasts (Bonferroni adjusted  $p$  value < 0.05) (Fig. 3a, dark grey points). Of these, 156 probes had a 1.5-fold difference in gene expression between airway and parenchymal fibroblasts, with 111 genes being more highly expressed in airway fibroblasts (Fig. 3a, blue points) and 45 more highly expressed in parenchymal fibroblasts (Fig. 3a, red points).

Between airway and parenchymal fibroblasts, 42 genes showed differential gene expression associated with differential methylation of at least one DNA methylation probe (104 CpG probes total) (gene names and CpG sites are provided in Supplemental Table 3). Gene set enrichment analysis of the 42 genes identified significant enrichment (enrichment score 1.7) for extracellular matrix constituents (Supplemental Table 4), suggesting DNA methylation may differentially underlie fundamental matrix deposition function in these two cell types. We found both direct and inverse relationships between differences in DNA methylation and fold changes in gene expression (Fig. 3b). Of the 104 differentially DNA methylated CpGs associated with differential gene expression, 68 probes had an inverse relationship, while 36 showed a direct relationship. The direction of the relationship was not associated with gene feature positioning (inverse relationship  $\chi^2$   $p$  value = 0.9287, direct relationship  $\chi^2$   $p$  value < 0.3005) (Fig. 3c) or CpG density (inverse relationship  $\chi^2$   $p$  value = 0.1769, direct relationship  $\chi^2$   $p$  value < 0.0610) (Fig. 3d).

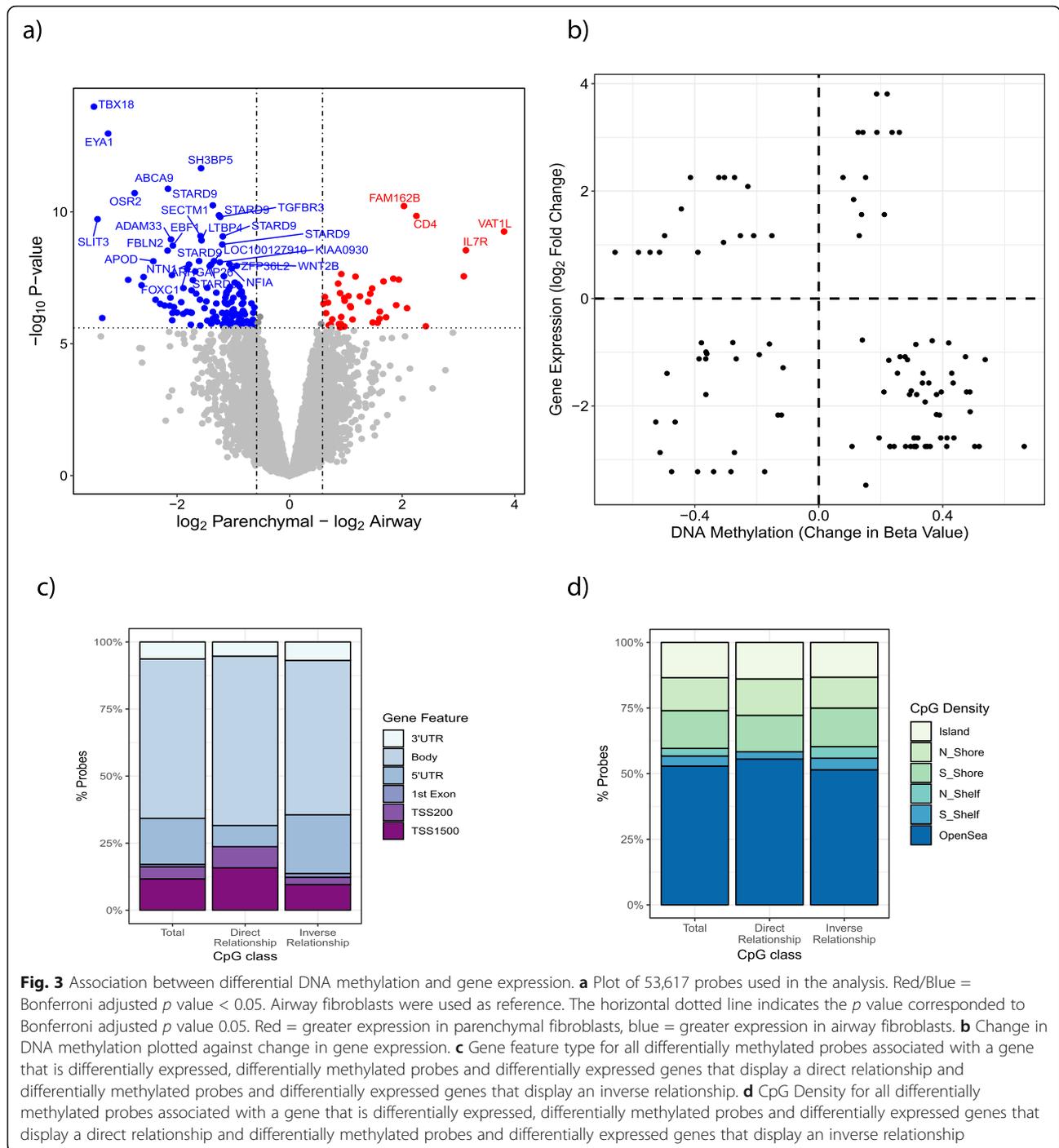
Genomic location, CpG density, delta beta (DNA methylation difference) values and the number of differentially methylated CpG sites associated with a particular gene are often used to infer biological relevance to DNA methylation data in the absence of gene expression data. To test whether these CpG methylation factors influenced gene expression differences, we compared all differentially methylated CpGs (3936 CpGs) that were annotated to a gene, to the differentially methylated

CpGs associated with the 158 genes differentially expressed between airway and parenchymal fibroblasts. We did not find any enrichment for genomic location (Fig. 4a,  $\chi^2$   $p$  value = 0.4442), density level (Fig. 4b,  $\chi^2$   $p$  value = 0.9235), delta beta (Fig. 4c, Kolmogorov–Smirnov  $p$  = 0.351) or number of differentially methylated CpGs (Fig. 4d, Kolmogorov–Smirnov  $p$  = 0.137). Specifically, for the 42 genes differentially expressed between airway and parenchymal fibroblasts which mapped to a CpG site with differential DNA methylation, we showed that despite DNA methylation beta values ranging from 7.8 to 50% and being associated with a range of < 1–2.25 log<sub>2</sub> fold differences in gene expression, there was no correlation (Fig. 4e,  $p$  = 0.1075). However, there was a statistically significant but weak correlation, between the number of differentially methylated CpG sites associated with a gene and the extent of differential gene expression (Fig. 4f, Spearman  $r$  = 0.457,  $p$  value = 0.00233). These data highlight that DNA methylation should be understood in association with gene expression rather than as a proxy, but that the number of differentially methylated CpGs potentially infers more relevance to gene expression effect size than CpG genomic location, delta beta or density.

#### Airway and parenchymal fibroblast DNA methylation profiles are differentially associated with asthmatic status

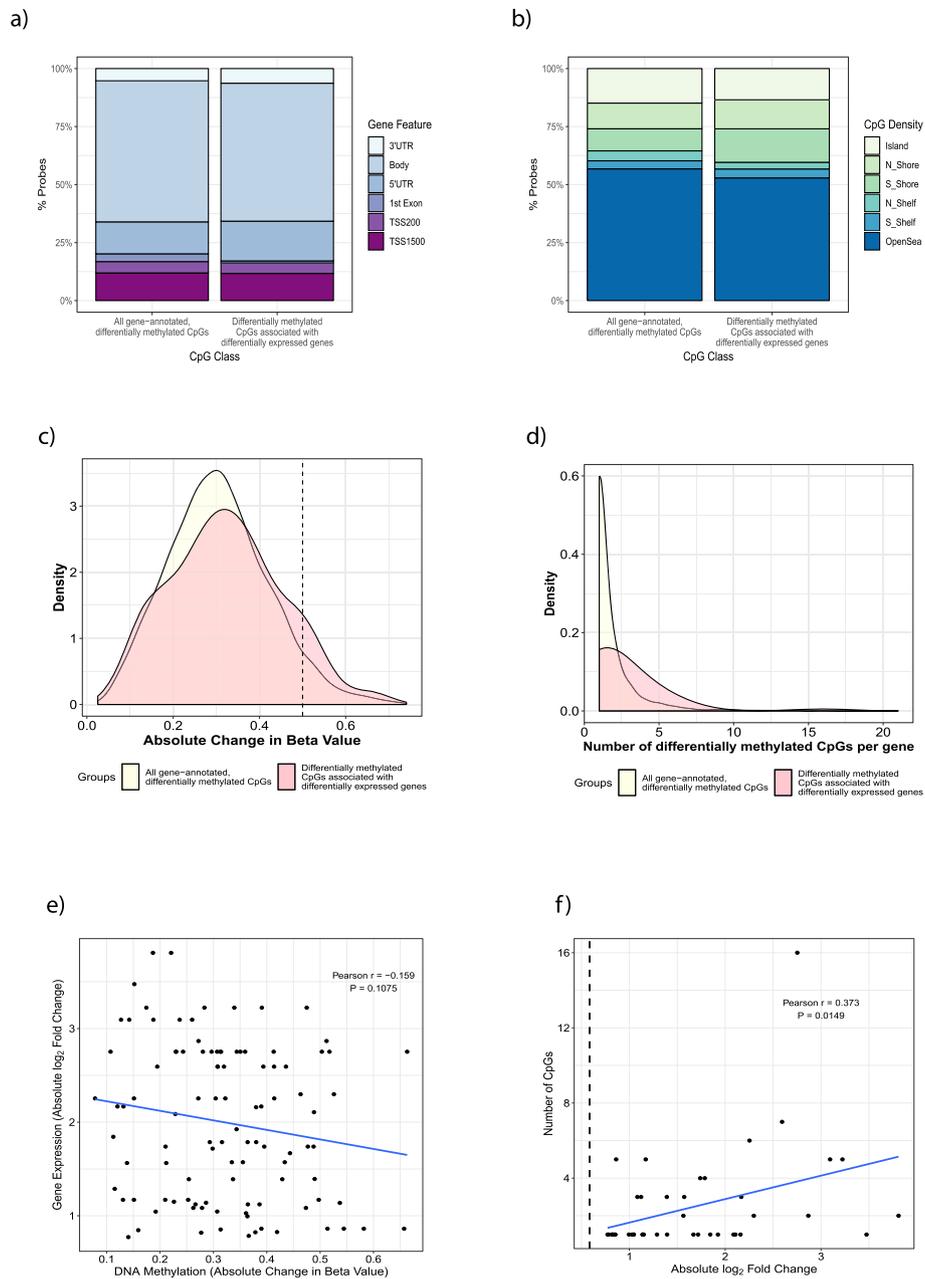
Having further identified molecular distinction between airway and parenchymal fibroblasts, we investigated the disease relevance of this distinction by assessing whether fibroblast DNA methylation profiles differed with asthmatic status, and importantly whether modifications were shared or distinct between airway and parenchymal fibroblasts. The demographics of the individuals from which cells were isolated are shown in Table 2.

To test whether asthma was associated with differential methylation, we performed a linear regression and did not observe any statistical differences in DNA methylation at individual CpG sites between asthma and non-asthma cases in either airway or parenchymal lung fibroblasts (Benjamini-Hochberg FDR < 0.05). However, a deviation of the raw  $p$  value distribution from the null hypothesis suggested an association between asthmatic status and DNA methylation in parenchymal fibroblasts (Fig. 5a, Kolmogorov–Smirnov  $p$  = 2.2e–16), but not airway fibroblasts (Fig. 5a, Kolmogorov–Smirnov  $p$  = 0.999). To further investigate this, we looked at aggregated sites via a regional analysis using R package DMRcate. This identified 17 genomic regions differentially methylated in association with asthmatic status in airway fibroblasts (Fig. 5b and details in Supplemental Table 5), and 112 regions in parenchymal fibroblasts (Fig. 5c



and details in Supplemental Table 6), with no overlap between the two cell types. The absence of overlap between asthma-associated DNA methylation differences between airway and parenchymal fibroblasts highlights the necessity for them to be investigated independently in lung disease studies, rather than considering a single lung fibroblast population. In airway fibroblasts, 15 of the 17 regions contained three

or more CpG sites, of which, 11 were annotated to a known gene promoter (reference genome hg19). Six had a maximum difference in DNA methylation (i.e. at least one probe displayed a mean difference in methylation) of greater than 20% ( $\Delta\beta = 0.2$ ) and were associated with the following genes: *PODN*, *HOOK2*, *RP11-214O1.2*, *HOXA7*, *HNF1A/HNF1A-AS1* and *C5orf38/IRX2* (Fig. 6a–f respectively). In parenchymal



**Fig. 4** Relationship between differential DNA methylation and gene expression. **a** Gene feature type for all differentially methylated probes and differently methylated probes associated with a gene that is differentially expressed. **b** CpG Density for all differentially methylated probes and differently methylated probes associated with a gene that is differentially expressed. **c** Distribution of CpG probe methylation difference in all differentially methylated probes (yellow) and differently methylated probes associated with a gene that is differentially expressed (pink). **d** Distribution of the number of CpG probes associated with all gene annotated, differentially methylated CpGs (Yellow) and of the number of differentially methylated probes associated with a gene that is differentially expressed (pink). **e** Correlation of gene expression and associated probe averaged DNA methylation changes. **f** Correlation of gene expression with number of differentially methylated CpG probes associated with the gene

fibroblasts, 99 of the 112 regions contained three or more CpG sites, of which 71 were annotated to a known gene. Six of these had a maximum difference in DNA methylation of 20% and were associated with the following genes: *HRNR*, *NR2F1-AS1*, *GDNF*,

*HOXA5/6/HOXA-AS3*, *RBPI* and *HLA-F* (Fig. 7a–f respectively).

Gene expression of the six airway and six parenchymal fibroblast regions specified above identified no expression of *PODN*, *HNF1A/HNF1A-AS1*, *C5orf38*,

**Table 2** Asthmatic donors demographics

	Airway fibroblasts			Parenchymal fibroblasts		
	Non-asthmatic	Asthmatic	P value	Non-asthmatic	Asthmatic	p value
Total N	8	9		9	8	
Sex (M/F)	5/3	3/6	0.6193	7/2	3/5	0.1534
Age (mean ± SEM)	33.5 ± 8.166	17.67 ± 2.95	0.0754	27.11 ± 5.978	18.75 ± 3.06	0.2566
Smoking status (non/ex/current)	4/2/2	7/0/2	0.2505	5/1/3	6/0/2	0.5389
Bronchodilator (Y/N)	NA	6/3		N/A	5/3	
Inhaled steroid (Y/N)	N/A	1/8		N/A	1/7	
Oral steroid	N/A	1/8		N/A	1/7	
Fatal asthma (Y/N)	N/A	6/3		N/A	8/3	

*IRX2*, *HLA-F*, *HRNR* or *GDNF* in either fibroblast type. For the remaining genes (airway fibroblasts: *HOOK2* and *HOXA7* (Fig. 8a and b); parenchymal fibroblasts: *HOXA5*, *HOXA6*, *HOXA-AS3*, *RBI* and *NR2F1-AS1* (Fig. 9a–e)), there was no differential gene expression between donors with and without asthma.

#### Validation of TWIST DNA methylation as a molecular marker of airway and parenchymal fibroblasts

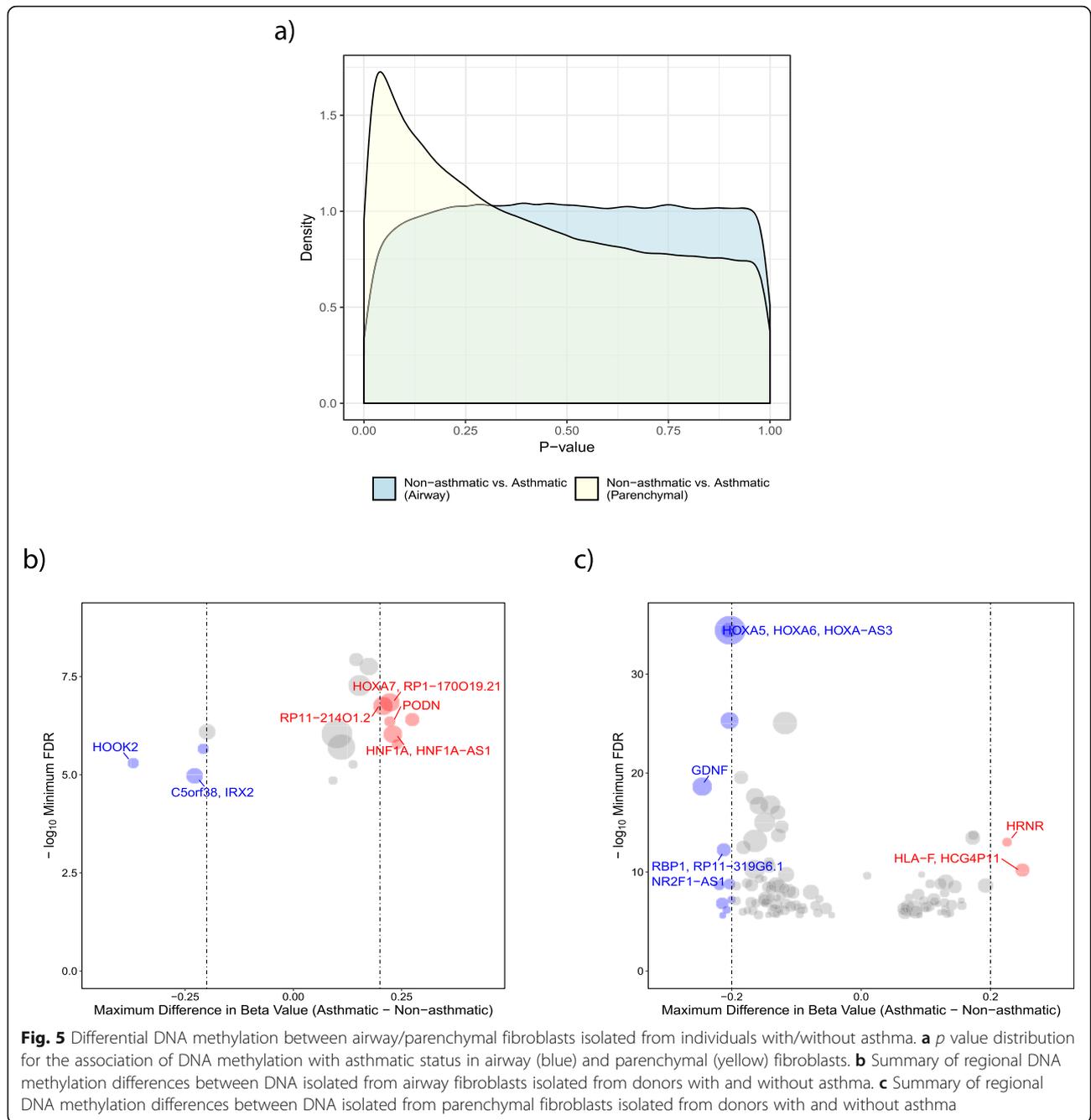
To investigate if DNA methylation could be used as a molecular cell-type marker, a validation cohort from seven donors with matched airway and parenchymal fibroblasts were used to remove any bias in DNA methylation levels due to inter-individual genetics and environmental exposures. Donors had a mean age of 23.42 (SEM ± 7.5), three were female, two current smokers and five ex-smokers. Eighty-eight of the 240 CpGs previously identified passed Bonferroni correction ( $p < 0.05$ ), and 78 of these exhibited an absolute beta difference of  $> 0.5$  (Fig. 10a). Details of these 78 CpGs are given in Supplemental Table 7. Of the significant CpGs, 67 with a beta difference of  $> 0.5$  were less methylated in parenchymal fibroblasts compared to airway fibroblasts, while 11 CpGs were more methylated (Fig. 10a). To identify a marker with maximum signal-to-noise ratio, specificity and sensitivity, we utilized ‘DMRcate’, a Bioconductor R package [14], to identify de novo differentially methylated regions with CpGs in close genomic proximity. Fourteen regions containing three or more probes were identified as differentially methylated between airway and parenchymal fibroblasts (Fig. 10b and Supplemental Table 8), of which eight were annotated to a gene, allowing for assessment of both gene expression and DNA methylation as a distinguishing marker. This was further restricted to regions containing a CpG identified in the site-by-site analysis by linear modelling, which provided gene regions for further assessment: *TWIST1* (Fig. 10c), *HLX* (Fig. 10d) and *SKAP2* (Fig. 10e).

Affymetrix array data identified differential gene expression of *TWIST1* (Fig. 10f), *HLX* (Fig. 10g) and *SKAP2* (Fig. 10h) between airway and parenchymal fibroblasts, providing three high confidence targets for differentiating between airway and parenchymal fibroblasts. Pyrosequencing verified differential DNA methylation of *TWIST1* (Fig. 11a), *HLX* (Fig. 11b) and *SKAP2* (Fig. 11c) between airway and parenchymal fibroblasts; however, qPCR only validated differential gene expression of *TWIST1* (paired samples Fig. 11d–f, full sample set Supplemental Fig. 1). Further, *TWIST1* protein was differentially expressed (higher expression in airway fibroblasts,  $p < 0.05$  two sample *t* test), strengthening the evidence for *TWIST1* as a DNA methylation cell-type marker capable of distinguishing airway and parenchymal fibroblasts (Fig. 11g/h).

Finally, logistic regression showed gene expression of *TWIST1* identified airway fibroblasts from parenchymal fibroblasts (Fig. 12a) at a validation area under the receiver operating characteristic curve of 92% (95% confidence interval 73.7–100%) (Fig. 12b), while elastic net regularized logistic regression showed DNA methylation of the six *TWIST1* CpGs fully distinguished airway fibroblasts from parenchymal fibroblasts (Fig. 12c) at a validation area under the receiver operating characteristic curve of 100% (Fig. 12d). This further suggested that the DNA methylation of CpG sites associated with the *TWIST1* gene can be utilized as a molecular marker to distinguish airway and parenchymal fibroblasts for future research.

#### Discussion

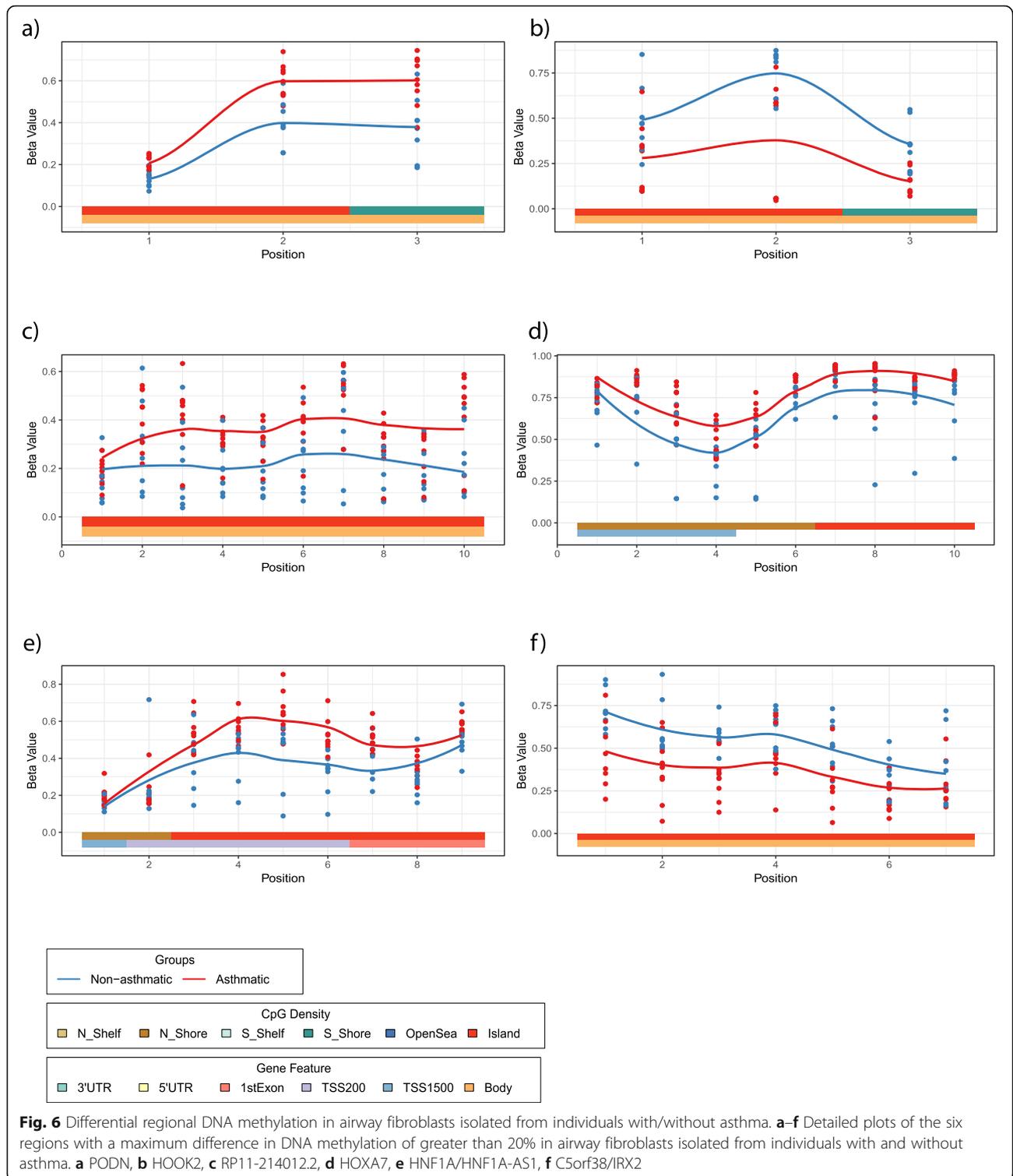
In the current study, we report that DNA methylation can be utilized as a cellular marker to distinguish airway and parenchymal lung fibroblasts using *TWIST1*; further, that DNA methylation profiles can identify differences between airway and parenchymal fibroblasts from asthmatic and non-asthmatic subjects; and lastly, that



the relationship between DNA methylation profiles and gene expression signatures is complex.

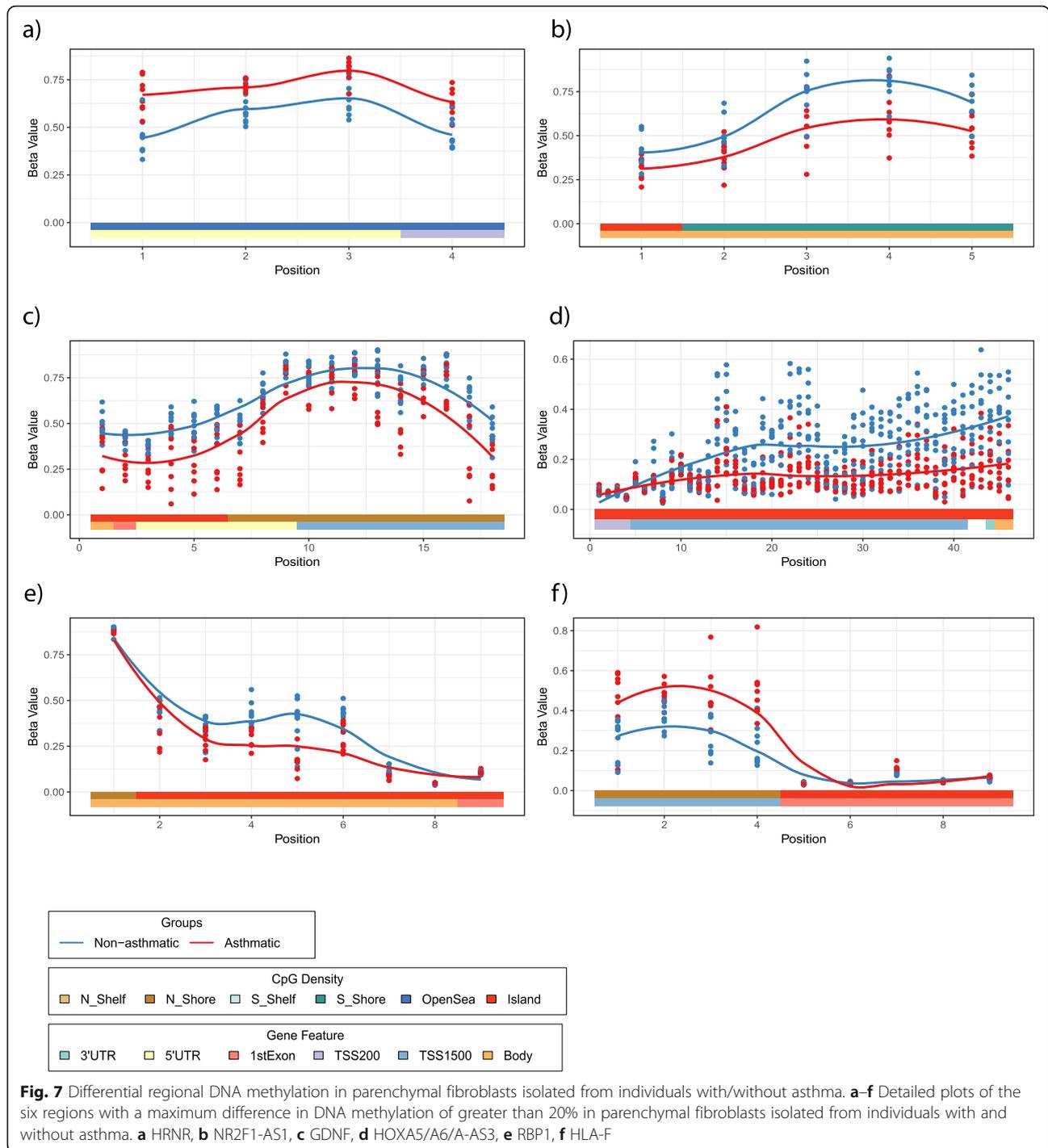
Airway and parenchymal lung fibroblast transcriptomes have previously been profiled [5], and our data confirmed that the two cell types do indeed have differential gene expression profiles. However, only 123 of the 963 differentially expressed genes in our study (1.5-fold change) replicated the 775 genes previously identified for airway and parenchymal lung fibroblasts by Zhou et al. [5]. One reason could be the use of a more stringent fold change cut off in our analysis (> 1.5 vs > 1.2);

however, a sub-analysis of our data showed relaxing our fold-change cut off to that of the previous study still only identified 183 of the previously reported genes. Zhou et al. also combined fibroblasts from asthmatic and non-asthmatic donors in their analysis, having first found no statistical differences in gene expression between the diseased and non-diseased samples. Lastly, technical factors including the cell passage number studied (p4 vs p3 in the Zhou study), the utilization of different microarray platforms, the different normalization methods (Robust Multiarray Averaging (RMA) vs cyclic



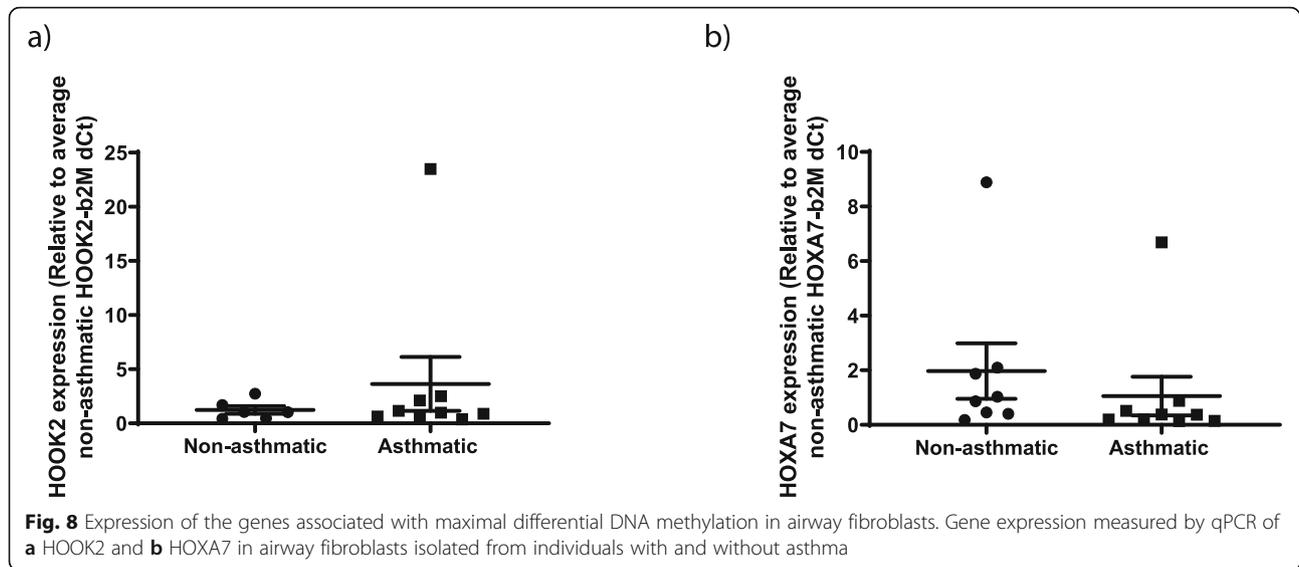
loss in the Zhou study) or the different statistical tests (limma moderated t-test versus Wilcoxon signed-rank test in the Zhou study) may also account for the discrepancies in our results as compared to the Zhou study. Regardless, these data highlight the transient and variable

nature of gene expression which reduces its ability to effectively distinguish between the cell types. Conversely, DNA methylation is a more stable molecular mark, and tissue-specific differentially methylated regions (tDMRs) have been identified in other tissues [9]



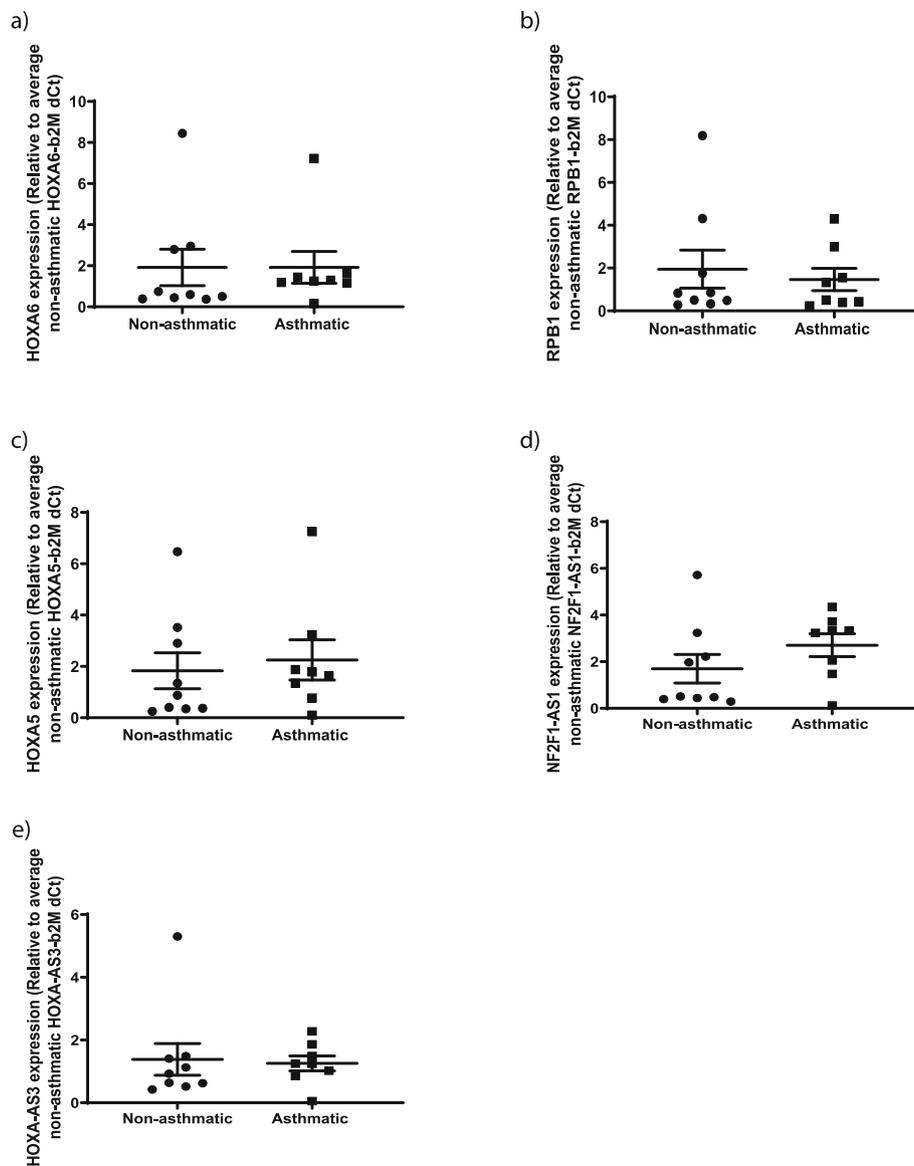
suggesting increased potential for utilization as a cellular marker. This study identified 3936 CpGs that were differentially methylated between airway and parenchymal fibroblasts in a cohort of unmatched airway and parenchymal fibroblasts. The airway and parenchymal fibroblast-specific CpGs identified were all enriched within gene bodies and open sea regions. This finding is corroborated by previously identified tissue-specific

differentially methylated regions (tDMRs) in blood, saliva, buccal swabs, hair follicles, liver, muscle, pancreas, subcutaneous fat, omentum and spleen that have been shown to be enriched in CpG-poor regions [9]. Interestingly, only 104 of our 3936 (2.6%) differentially methylated CpGs were associated with differential gene expression when using the closest annotated gene name to define the CpG/gene association. Further, the link



between gene expression and DNA methylation was not explained by CpG location (gene feature or CpG density), the size of the change in methylation level (delta beta), or the number of differentially methylated CpGs associated with that gene. However, there was a weak correlation between the number of differentially methylated CpGs associated with a gene and the extent of the difference in gene expression. These findings highlight that the relationship between DNA methylation and gene expression differences is complex and need to be validated. As there was no available datasets for us to corroborate our initial finding, we validated the identified 240 high confidence CpGs in a second cohort of matched airway and parenchymal fibroblasts and further confirmed TWIST1 as a cell marker for airway versus parenchymal lung fibroblasts at the DNA methylation, gene expression and protein level. Importantly, elastic net-regularized logistic regression analysis demonstrated that the DNA methylation profile of TWIST1 provided 100% separation between the two cell types indicating that the larger and more defined differences in DNA methylation perform better as a cell-type marker. TWIST1 is a 21 kDa transcription factor [15] that is primarily expressed in mesoderm tissues from the early stages of embryo development and is involved in the specification and differentiation of mesenchyme tissues. It is therefore probable that TWIST1 is a cell marker for airway and parenchymal fibroblasts due to its distinct regulation of mesenchymal cell phenotypes. The lack of an independent validation cohort remains a limitation to the study; however, the cells used in the current study were isolated at three different geographical locations using the same tissue culture protocols, limiting the opportunity for our differences to be due to isolation bias.

Understanding the regulation of genes like TWIST1 by DNA methylation may have important implications for understanding the role of airway and parenchymal fibroblasts in lung health and disease. Differential DNA methylation in asthma has been studied in the airway epithelium [16–19], but not in lung fibroblast populations. We identified 17 airway and 112 parenchymal fibroblast differentially methylated DNA regions that were associated with asthma, with no overlap between the two cell types. None of the airway but two of the parenchymal fibroblast asthma-associated DNA methylation regions were annotated to genes with previously identified genome-wide association study genetic risk loci for asthma [20], TSLP (rs1837253) and GATA3 (rs2589561), with the SNP and the differentially methylated regions being between ~ 0.6 and 1 Mbp apart. Importantly, the differences in the number of asthma-associated differences in DNA methylation observed between airway and parenchymal fibroblasts was independent of fatal and non-fatal asthma, and the therapeutic use of inhaled bronchodilators and steroids as airway and parenchymal fibroblasts were isolated from the same asthmatic individuals. The larger number of perturbations to DNA methylation in association with asthma in parenchymal fibroblasts versus airway fibroblasts was surprising based on the understanding that the parenchyma's contribution to the asthmatic phenotype is thought to be minimal. However, airway fibrosis is driven by myofibroblasts (an  $\alpha$  smooth muscle actin ( $\alpha$ SMA) positive fibroblast subtype), and recently an  $\alpha$ SMA-positive fibroblast subtype was identified in the lung parenchyma, with approximately three times more  $\alpha$ SMA-positive cells in the parenchyma of individuals with asthma compared to non-asthmatic control subjects [21, 22] suggesting that parenchymal-derived fibroblasts in asthmatics may play a role. Furthermore, it has been shown that there is more



**Fig. 9** Expression of the genes associated with maximal differential DNA methylation in parenchymal fibroblasts. Gene expression measured by qPCR of **a** HOXA6, **b** RPB1, **c** HOXA5, **d** NF2F1-AS1 and **e** HOXA-AS3 in parenchymal fibroblasts isolated from individuals with and without asthma

parenchymal extracellular matrix in asthmatic lungs compared to controls [22]. These studies highlight an emerging role for the parenchyma fibroblast in asthma pathology and the necessity to work with the most appropriate cell type when investigating human disease.

## Conclusions

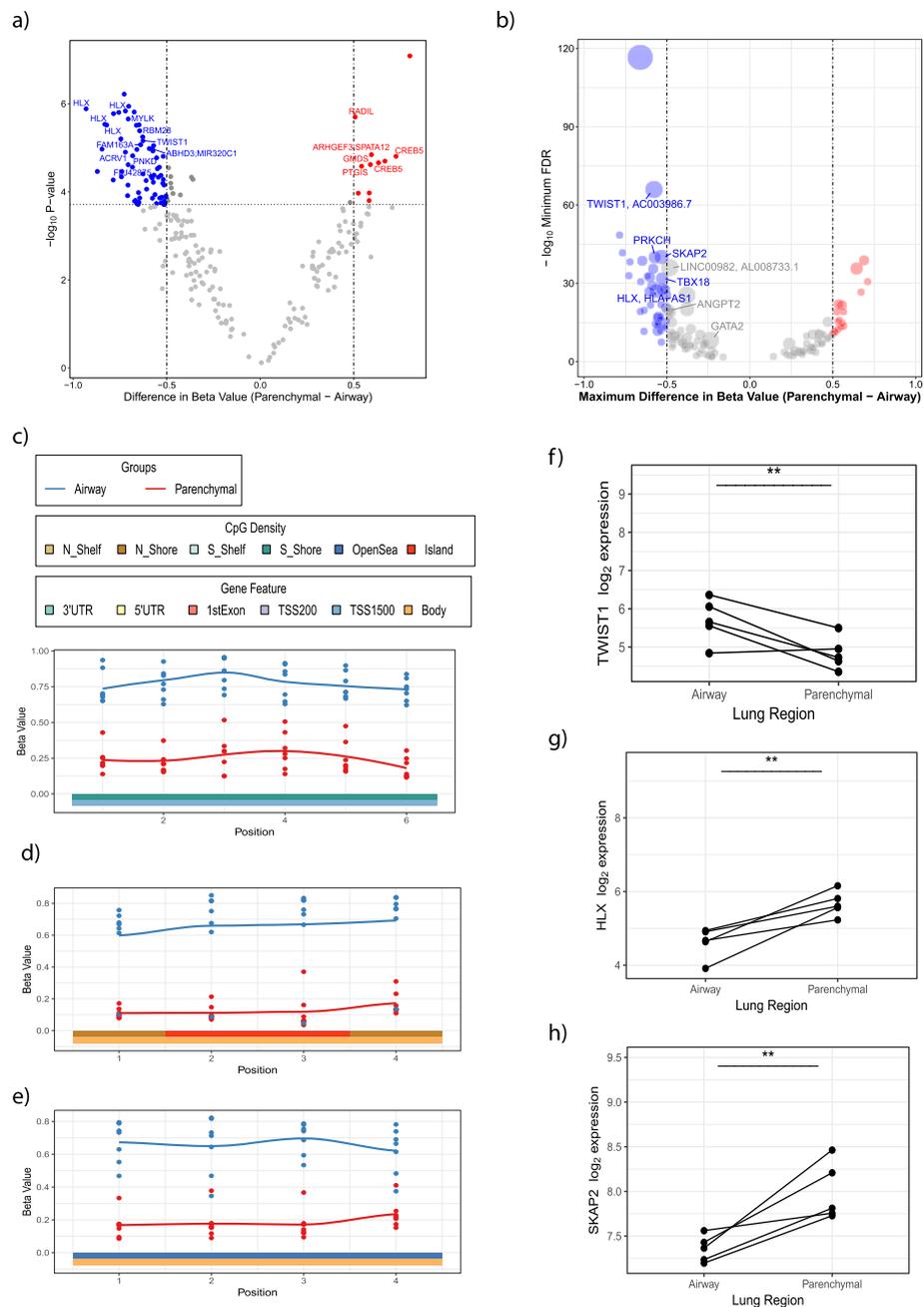
In conclusion, our study identified in two independent sample cohorts, that genome-wide and targeted DNA methylation profiles of *TWIST1* can distinguish between airway and parenchyma-derived lung fibroblasts. Further, airway and parenchymal fibroblast-related differences in DNA methylation and associated gene expression, indicate

these two cells are phenotypically different and likely perform separate and distinct roles in normal lung physiology. Further, we show that airway and parenchymal fibroblast DNA methylation is differentially altered in individuals with asthma and the role of both cell types should be considered in the pathogenesis of asthma.

## Methods

### Isolation and culture of airway and parenchymal fibroblasts

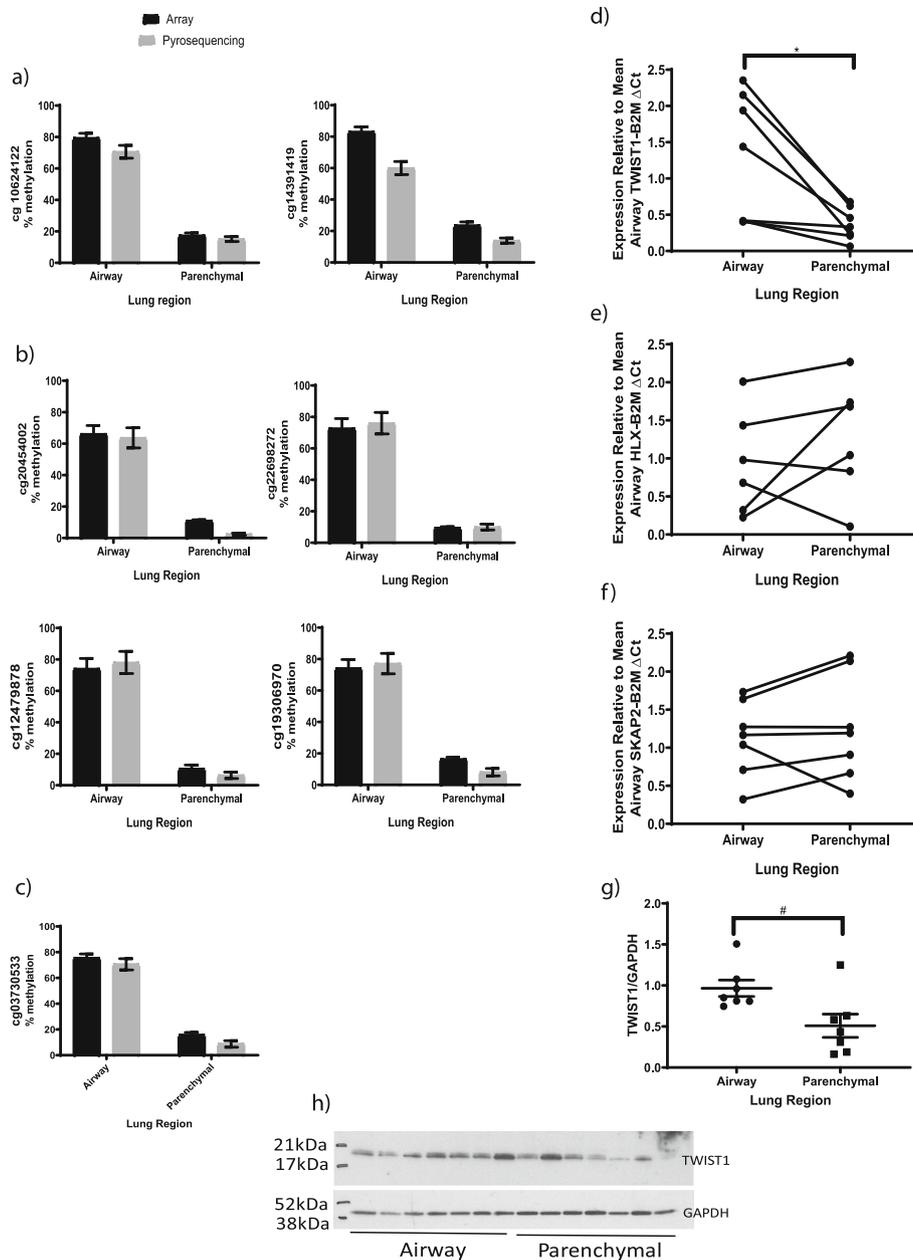
Primary cultures of airway and parenchymal fibroblasts from healthy and asthmatic individuals were isolated from lung biopsies, intrapulmonary airway and



**Fig. 10** Validation of differential methylation between airway and parenchymal fibroblasts and associated gene expression. **a** Plot of 240 CpG probes used in the analysis. Red/Blue = Bonferroni adjusted  $p$  value  $< 0.05$ . Red = greater expression in parenchymal fibroblasts, blue = greater expression in airway fibroblasts. **b** Summary of regional DNA methylation differences within the 240 CpG probes used in the analysis, between DNA isolated from airway versus parenchymal fibroblasts. **c** Regional beta value plots for TWIST1 in paired samples. **d** Regional beta value plot for HLX in paired samples. **e** Regional beta value plot for SKAP2 in paired samples. Microarray generated gene expression in paired airway and parenchymal fibroblasts for **f** TWIST1, **g** HLX and **h** SKAP2 (\*\*limma moderated  $t$  test  $p < 0.01$ )

parenchymal lung tissue obtained from lung cancer resections (from the normal margin) or non-transplantable donor lungs. Airway and parenchymal fibroblasts were derived using the outgrowth techniques as previously described [3, 23] at three different sites. Briefly, 2 mm<sup>2</sup> tissue explants were placed in 6-well tissue culture plates

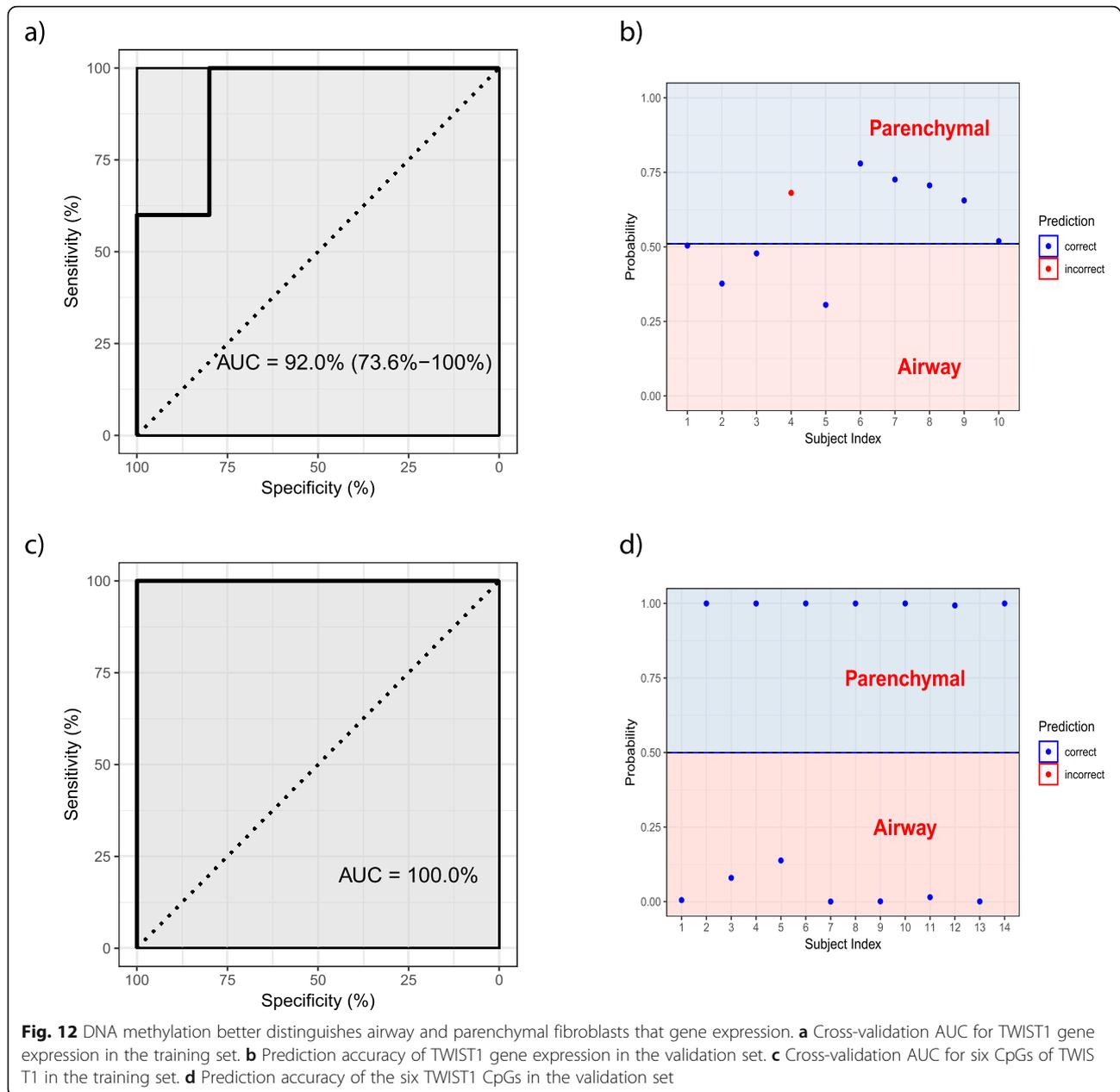
with DMEM (Sigma) containing 10% fetal bovine serum (GIBCO, Life Technologies), penicillin (100 U/ml), streptomycin (100 µg/ml) and L-glutamine (4 mM) in a 5% CO<sub>2</sub>-humidified incubator. Media was replaced regularly until cellular outgrowth reached confluence. Tissue pieces were removed and destroyed, and cells were



**Fig. 11** Validation of high confidence targets. Pyrosequencing versus array percent methylation for CpGs identified in regional CpG methylation analyses for **a** TWIST1, **b** HLX and **c** SKAP2. QPCR generated gene expression in the paired samples for **d** TWIST1, **e** HLX, **f** SKAP2 (\*Wilcoxon signed-rank test  $p < 0.05$ ). **g** Densitometry and **h** associated western blot for TWIST1 whole cell protein expression ( $^{\#}p < 0.05$  by two-sample  $t$  test,  $n = 7$  airway,  $n = 7$  parenchymal)

harvested using trypsin/EDTA solution (Sigma). All cells for this study were cultured at the same time under the exact same conditions. Samples were generated from cells at passage 4 except for two asthmatic airway fibroblast samples at passage 5, a single healthy airway fibroblasts sample at passage 3, and two asthmatic parenchymal fibroblast samples at passage 5. Cells at the required passage were grown to confluence in 6-well plates and serum starved for 24 h prior to lysis for DNA,

RNA and protein isolation. The tissue was obtained and cells extracted with the approval of each of the research ethics boards for each of the academic institutions involved: McMaster University (Hamilton Integrated Research Ethics Board Ref:00-1839), University of British Columbia (Providence Health Care Research Ethics Board Ref:H13-02173) and University of Nottingham (East Midlands Research Ethics Committee Ref: 08/H0407/1).



### DNA and RNA isolation

DNA and RNA were simultaneously isolated from each sample using the AllPrep DNA/RNA Mini Kit (Qiagen) as per the manufacturer's instructions and assessed for quality and quantity using a NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific).

### Protein isolation

Cells were washed with PBS, lysed in RIPA buffer (Sigma) and stored at  $-80^{\circ}\text{C}$  until required.

### Bisulfite conversion and DNA methylation arrays

Seven hundred fifty nanograms of purified genomic DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research) as per the manufacturer's instructions. Specific incubation conditions for the Illumina Infinium Methylation Assay were used as per the manufacturer's protocol Appendix. Samples were eluted in  $12\ \mu\text{l}$  of the provided elution buffer. Bisulfite-converted DNA was assessed for concentration and quality using a NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific), and 160 ng of the conversion product was used for genome-wide DNA methylation

quantification at over 485,000 CpG sites using the Illumina Infinium HumanMethylation450 BeadChip array, according to the manufacturer's protocols.

#### DNA methylation data quality control and normalization

IDAT files produced by GenomeStudio were imported into the R statistical software (version 3.2.1) using the *minfi* package (v. 1.14.0) [24]. The 65 known quality control SNP probes were used to cluster all samples to detect anomalies within samples from the same donor. Probes were excluded from further analysis according to several criteria: first, 1402 probes were found to have either a detection  $p$  value  $< 0.05$  in at least 1% of samples or had less than 3 bead count in at least 5% of samples; second, the 65 SNP probes; third, 59,593 probes were found to be cross-hybridized to other parts of the genome [25]; and fourth, 9925 probes on the XY chromosomes. 414,592 probes remained for analysis. Filtered probes were normalized using the *funtooNorm* algorithm [26], which extends the *funNorm* procedure [27] and is purported to correct for unwanted variation while preserving important differences in methylation patterns between different cell types. We employed the normalization option of principal components regression with 5 principal components. Two values of DNA methylation were calculated, beta-values ( $\beta$ -values) and  $M$ -values.  $\beta$ -values are the ratio of all methylated probe intensities over total signal intensities (methylated and unmethylated) and have a range from 0 to 1. They approximately represent percent methylation.  $M$ -values are the logit transformation of  $\beta$ -values and are more statistically robust [28]. All statistical analyses were performed using  $M$ -values, while  $\beta$ -values were used for visualization and interpretability purposes. Principal components analysis was performed for quality control of the  $M$ -values. Two replicates between passage 3 and passage 4 of the same donor provided correlation  $r^2$  of 0.9869 and 0.9948, suggesting minimal genome-wide DNA methylation dysfunction due to passage.

#### Differential DNA methylation analysis

##### *Airway versus parenchymal in healthy individuals*

Samples were split into two groups: those used in the initial analysis in which we used all available samples regardless of whether matched samples from airway and parenchyma were available, and a second group, containing only samples from donors from which we obtained both airway and parenchymal samples. The paired samples were not included in the initial analysis and represent a completely independent data set.

Linear regression analysis was applied to the initial group using the *limma* package in R [29] while adjusting for age as a covariate. As large changes in DNA methylation were apparent, we considered only CpG sites to be

significant if they had a Bonferroni-adjusted  $p$  value  $< 0.05$ , and an effect size on methylation- $\beta$  of more than 0.5. With these sites, a similar analysis was done on the paired group using their DNA methylation differences between parenchyma and airway, which is analogous to a paired  $t$  test, but also adjusted for age. Filtering the latter results to chromosomes with more than one hit, we identified genomic clustering of the CpG sites in differentially methylated regions (DMRs) using the *DMRcate* package in R [14], which uses Gaussian kernel smoothing to find patterns of differential methylation, agnostic to genomic annotation. We used the authors' recommended bandwidth ( $\lambda$ ) of 1000 base pairs and scaling factor ( $C$ ) of 2, though we were using a sparse set of sites, the intention was to find regions with large effect sizes. Gene set testing was performed with the *methylglm* function of *methylGSA* [30] in R, which adjusts for the number of CpGs associated with a gene.

##### *Asthma versus non-asthma*

Airway and parenchymal fibroblasts were considered separately. Linear regression analysis using the *limma* package in R was used to identify individually differentially methylated CpG sites. Regional differences in DNA methylation associated with asthmatic status were identified using the *DMRcate* package in R including all CpGs specified by a nominal linear modelling  $p$  value limit of  $< 0.001$ .

#### Gene expression microarray analysis

Whole-genome transcriptome analysis was conducted by hybridizing samples of total RNA to Affymetrix Human Gene 2.1 ST Arrays Strips (Affymetrix, Santa Clara, CA, USA). A minimum RIN score of 8 was used as cut off for inclusion in the microarray analysis. Two hundred fifty-nanogram RNA was used for all samples. All steps were conducted at the Nottingham Arabidopsis Stock Centre.

#### Differential gene expression analysis

Raw CEL files were read into R, RMA background corrected with quantile normalization and log2 transformed using the *Oligo* package [31–34]. Linear regression analysis was applied to sample groups using the *limma* package in R to identify differentially expressed genes [29]. Both Benjamini-Hochberg and Bonferroni-corrected significance levels are reported. Integration with DNA methylation was done with Bonferroni-corrected data for both DNA methylation (3936 CpGs) and gene expression (158 genes), using Affymetrix gene annotation for the gene expression data and Illumina Closest Transcription Start Site annotation for DNA methylation data.

### Bisulfite PCR-pyrosequencing

DNA methylation data were confirmed using Pyrosequencing at 7 CpG sites present in both the DMRcate and linear modelling data comparing airway and parenchymal fibroblasts: TWIST1, cg10624122, cg14391419; HLX, cg20454002, cg22698272, cg12479878, cg19306970; SKAP2, cg03730533. We were unable to design a functional assay for SKAP2 cg12140851. Bisulfite PCR-pyrosequencing assays were designed with PyroMark Assay Design 2.0 (Qiagen). The regions of interest were amplified by PCR using the HotstarTaq DNA polymerase kit (Qiagen) as follows: 15 min at 95 °C (to activate the Taq polymerase), 45 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s, and a 5-min 72 °C extension step. For pyrosequencing, single-stranded DNA was prepared from the PCR product with the Pyromark™ Vacuum Prep Workstation (Qiagen) and sequencing was performed using sequencing primers on a Pyromark™ Q24 pyrosequencer (Qiagen). The quantitative levels of methylation for each CpG dinucleotide were calculated with Pyromark Q24 software (Qiagen). Primer sequences are shown in Table 3.

### Reverse transcription and qPCR

0.5 µg of RNA was reverse transcribed using SuperScript IV (Invitrogen) as per the manufacturer's instructions. The resulting 20 µl cDNA samples were diluted to a total volume of 200 µl using nuclease-free water. cDNA was amplified using PerfeCTa SYBR Green FastMix (Quanta bio), with 2 µl template and 200 nM primers in a 10-µl reaction using a Stratagene Mx3000P/3005P system. Thermal cycler conditions included incubation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 20 s. Data was collected in MxPro, a single product was confirmed by melt curve analysis and Ct values were exported to excel for analysis. Expression was expressed by the  $\Delta\Delta C_t$  method relative to  $\beta_2$ -Microglobulin ( $\beta_2M$ ) Ct and mean airway fibroblast target/ $\beta_2M$   $\Delta C_t$ . Primer sequences are as

follows:  $\beta_2$ -Microglobulin, forward 5'-AATCCAAATG CGGCATCT-3', reverse 5'-GAGTATGCCTGCCGTG TG-3'; TWIST1, forward 5'-GCCCCGAGACCTAGAT GTCATT-3', reverse 5'-CCCACGCCCTGTTTCT TTGA-3'; HLX, forward 5'-CGCTGAGAGATCTCAC TTCCC-3', reverse 5'-TCAGGATTGCAGAAGCCT CG-3'; SKAP2, forward 5'-GTTCTTAATCCGGGCC GCTA-3', reverse 5'-TCAACATCTGCCAACAGG TTC-3'. Primer utility was tested on reference cDNA (Takara) for positive amplification prior to assessment of samples.

### Western blotting

Fifteen-microgram whole cell lysate protein samples were subject to electrophoresis in 15% SDS-polyacrylamide gel. Separated proteins were electroblotted to polyvinylidene difluoride membranes, and the blot was blocked for 1 h at room temperature with blocking buffer (0.1% TBST with 5% fat-free dried milk powder). The blot was then incubated with 1:1000 dilution TWIST 1 (clone MAB6230, R&D Systems) or 1:10,000 dilution GAPDH (Abcam) at 4 °C overnight. The blot was washed with 0.1% TBST and incubated with HRP-conjugated anti-mouse secondary Abs (DakoCytomation, Cambridge, U.K.) (1:2000 dilution with 5% fat-free dried milk in 0.1% TBST) for 1 hour. The blot was washed again and then incubated with Clarity western ECL substrate (Bio-Rad). The densitometry analysis was performed in ImageJ.

### Classification using elastic net-regularized logistic regression

Elastic net-regularized logistic regression was applied to both DNA methylation data (microarray) and gene expression data (microarray) to distinguish airway fibroblasts from parenchymal fibroblasts. Data were split into a training set and a validation set. Samples from the initial non-matched groups were used as the training set while samples from the paired groups were used as the

**Table 3** Pyrosequencing primer details

Target	Forward primer	Reverse primer	Sequencing primer	Assay position
TWIST cg10624122	AGTATATAGTGTTGGGGTGGG	*AGTATATAGTGTTGGGGTGGG	AGGTTAGTTTATGGGTTTTTATAG	5
TWIST1 cg14391419	AGTATATAGTGTTGGGGTGGG	*ACCTTACTCCAACCCAAAAA	TTGGGGTGGGGGTAG	1
HLX cg20454002	ATGGAAGTGAGGATATATAGGAATTTT	*TCTCTCTCCCAACAATTACCAA	AGGGATATATAGGAATTTT	2
HLX cg22698272	*GGTGTATTTTTAGGTTTGTAGT	CCCCTTAAACAATACTCTAAAATTTT TCC	ACAAAAACCTCCTTAATAAAAT CT	1
HLX cg12479878	TTTTAGGATTGAAGTTTTTGGGTTGTT	*CTACCCCTTTTCAAAAAAACCA	GGTGTAGTAATTTTATAATTGGG	1
HLX cg19306970	AGTTTATGTTGGGTGTTGGATAT	*CCCTTAAACCTAAATAACAACC	TTTGGGTGTTGGATATA	1
SKAP1 cg03730533	GTAGTAATTTGATAATAAGAAAAGTTA GT	*ACCTCCCACCTTCTCTCCC	TTTAGTTAGTTTTTAGAATTT TTT	1

\*Biotinylated

validation set. Elastic net performs both shrinkage of regression coefficients and feature selection to identify features (genes or CpG sites) that are highly discriminative with respect to the phenotypes (fibroblast type). The mixing percentage ( $\alpha$ ) and regularization shrinkage parameter ( $\lambda$ ) were tuned at different values to build biomarker panels of various sizes. Area under the receiver operating characteristic curve (AUC) obtained from leave-one-out cross-validation (LOOCV) of the training set was used as the evaluation metric to identify the best-performing panel. The optimal cut-off for predicted probabilities was determined using Youden's index method to maximize both sensitivity and specificity. The validation set was subsequently tested through the best-performing panel to calculate the validation AUC and prediction accuracy. All the analyses were performed using the *glmnet* package in R [35].

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13148-020-00931-4>.

**Additional file 1.** qPCR generated gene expression of paired airway and parenchymal fibroblasts for TWIST1, HLX and SKAP2.

**Additional file 2.** Supplementary Tables 1-8

### Abbreviations

ECM: Extracellular matrix; FSP-1: Fibroblast-specific protein 1; TGF $\beta$ 1: Transforming growth factor beta 1; SEM: Standard error of the mean; FDR: False discovery rate; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

### Acknowledgements

Not applicable

### Authors' contributions

RLC devised the study; cultured cells; isolated DNA; analysed array, qPCR and pyrosequencing data; and wrote the manuscript. YY analysed array data. NF analysed array data. JP generated qPCR, pyrosequencing data and western blot data. JLM and LMM designed DNA methylation microarray experiments and performed sample array processing. STM and MC designed gene expression microarray experiments and performed sample array processing. PN provided lung tissue. MO provided statistical expertise. MSK facilitated RLC's analysis of array data and contributed to manuscript preparation. AJK oversaw the project and contributed to the manuscript preparation. TLH helped devise the study, provided lung tissue, isolated cells, statisticians (NF, YY) for the analysis and contributed to the manuscript preparation. All authors read and approved the final manuscript.

### Funding

The listed funding body's had no role in the design, data collection, analysis, interpretation or writing of the manuscript. RLC was supported by the European Respiratory Society (ERS) and the Canadian Thoracic Society (CTS)/Canadian Lung Association, joint ERS/CTS Long-Term Research Fellowship LTRF2013, funding from Rosetrees Trust, The Henry Smith Charity and a current Fellowship from the Medical Research Foundation (MRF)/Asthma UK. Support for this study was provided by Canadian Institutes of Health Research (CIHR) and British Columbia Lung Association operating grants. T-LH was supported by CIHR, Michael Smith Health Research Foundation (MSHRF), and Parker B. Francis New Investigator awards. PN is supported by the Frederick E. Hargreave Teva Innovation Chair in Airway Diseases. MSK is the Canada Research Chair in Social Epigenetics, Senior Fellow of the Canadian Institute for Advanced Research, and Sunny Hill BC Leadership Chair in Child Development. LM McEwen is supported by a CIHR Doctoral Research Award

(F15-04283). AJK is associated with NIHR Nottingham Biomedical Research Centre and the Nottingham MRC Molecular pathology node.

### Availability of data and materials

The dataset generated and analysed during the current study is available in the Gene Expression Omnibus (GEO) repository, <https://www.ncbi.nlm.nih.gov/geo/> GSE157651.

### Ethics approval and consent to participate

The tissue and extracted cells were obtained with informed consent and the approval of each of the research ethics boards for each of the academic institutions involved: McMaster University (Hamilton Integrated Research Ethics Board Ref:00-1839), University of British Columbia (Providence Health Care Research Ethics Board Ref:H13-02173) and University of Nottingham (East Midlands Research Ethics Committee Ref: 08/H0407/1).

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Nottingham NIHR Biomedical Research Centre, Nottingham MRC Molecular Pathology Node, Division of Respiratory Medicine, University of Nottingham, Nottingham University Hospitals NHS Trust, City Hospital, Nottingham, UK. <sup>2</sup>Centre for Heart Lung Innovation, University of British Columbia, Vancouver, Canada. <sup>3</sup>BC Children's Hospital Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. <sup>4</sup>Nottingham Arabidopsis Stock Centre, Plant Sciences Building, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK. <sup>5</sup>Firestone Institute for Respiratory Health, St Joseph's Healthcare and Department of Medicine, McMaster University, Hamilton, Ontario, Canada. <sup>6</sup>Department of Anaesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, Canada.

Received: 22 April 2020 Accepted: 1 September 2020

Published online: 02 October 2020

### References

- Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet*. 2006;2:e119.
- Chang HY, Chi JT, Dudoit S, Bondre C, van de Rijn M, Botstein D, et al. Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci U S A*. 2002;99:12877–82.
- Pechkovsky DV, Hackett TL, An SS, Shaheen F, Murray LA, Knight DA. Human lung parenchyma but not proximal bronchi produces fibroblasts with enhanced TGF- $\beta$  signaling and  $\alpha$ -SMA expression. *Am J Respir Cell Mol Biol*. 2010;43:641–51.
- Dessalle K, Narayanan V, Kyoh S, Mogas A, Halayko AJ, Nair P, et al. Human bronchial and parenchymal fibroblasts display differences in basal inflammatory phenotype and response to IL-17A. *Clin Exp Allergy*. 2016;46:945–56.
- Zhou X, Wu W, Hu H, Milosevic J, Konishi K, Kaminski N, et al. Genomic differences distinguish the myofibroblast phenotype of distal lung fibroblasts from airway fibroblasts. *Am J Respir Cell Mol Biol*. 2011;45:1256–62.
- Chu HW, Halliday JL, Martin RJ, Leung DY, Szefer SJ, Wenzel SE. Collagen deposition in large airways may not differentiate severe asthma from milder forms of the disease. *Am J Respir Crit Care Med*. 1998;158:1936–44.
- Roche WR, Beasley R, Williams JH, Holgate ST. Subepithelial fibrosis in the bronchi of asthmatics. *Lancet*. 1989;1:520–4.
- Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*. 2008;18:1518–29.
- Sliker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450 k array. *Epigenetics Chromatin*. 2013;6:26.

10. Schilling E, Rehli M. Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics*. 2007;90:314–23.
11. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet*. 2007;3:2023–36.
12. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 2011;6:692–702.
13. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*. 2009;41:178–86.
14. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samarasinghe K, Lord RV, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015; 8:6.
15. Thisse B, el Messal M, Perrin-Schmitt F. The twist gene: isolation of a drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Res*. 1987;15:3439–53.
16. Stefanowicz D, Hackett TL, Garmaroudi FS, Gunther OP, Neumann S, Sutanto EN, et al. DNA methylation profiles of airway epithelial cells and PBMCs from healthy, atopic and asthmatic children. *PLoS One*. 2012;7:e44213.
17. Yang Y, Haitchi HM, Cakebread J, Sammut D, Harvey A, Powell RM, Holloway JW, Howarth P, Holgate ST, Davies DE. Epigenetic mechanisms silence a disintegrin and metalloprotease 33 expression in bronchial epithelial cells. *J Allergy Clin Immunol*. 2008; 121:1393–1399, 1399 e1391–1314.
18. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight*. 2016;1:e90151.
19. Xu CJ, Soderhall C, Bustamante M, Baiz N, Gruzjeva O, Gehring U, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med*. 2018;6:379–88.
20. Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy Asthma Immunol Res*. 2019;11:170–87.
21. Boser SR, Mauad T, Araujo-Paulino BB, Mitchell I, Shrestha G, Chiu A, et al. Myofibroblasts are increased in the lung parenchyma in asthma. *PLoS One*. 2017;12:e0182378.
22. Weitoft M, Andersson C, Andersson-Sjolund A, Tufvesson E, Bjermer L, Erjefalt J, et al. Controlled and uncontrolled asthma display distinct alveolar tissue matrix compositions. *Respir Res*. 2014;15:67.
23. Noordhoek JA, Postma DS, Chong LL, Menkema L, Kauffman HF, Timens W, et al. Different modulation of decorin production by lung fibroblasts from patients with mild and severe emphysema. *COPD*. 2005;2:17–25.
24. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363–9.
25. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin*. 2013;6:4.
26. Oros Klein K, Grinek S, Bernatsky S, Bouchard L, Ciampi A, Colmegna I, Hudson M, et al. funtooNorm: an R package for normalization of DNA methylation data when there are multiple cell or tissue types. *Bioinformatics*. 2016;32:593–5.
27. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450 k methylation array data improves replication in large cancer studies. *Genome Biol*. 2014;15:503.
28. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015.
30. Ren X, Kuan PF. methylGSA: a Bioconductor package and shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics*. 2019;35:1958–9.
31. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. 2010;26:2363–7.
32. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19:185–93.
33. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31:e15.
34. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249–64.
35. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

