



OPEN

# DNA methylation and gene expression signatures are associated with ataxia-telangiectasia phenotype

Sharon A. McGrath-Morrow<sup>1</sup>✉, Roland Ndeh<sup>1</sup>, Kathryn A. Helmin<sup>2</sup>, Basil Khuder<sup>2</sup>, Cynthia Rothblum-Oviatt<sup>3</sup>, Joseph M. Collaco<sup>1</sup>, Jennifer Wright<sup>4</sup>, Paul A. Reyfman<sup>2</sup>, Howard M. Lederman<sup>4</sup> & Benjamin D. Singer<sup>2,5,6</sup>

People with ataxia-telangiectasia (A-T) display phenotypic variability with regard to progression of immunodeficiency, sino-pulmonary disease, and neurologic decline. To determine the association between differential gene expression, epigenetic state, and phenotypic variation among people with A-T, we performed transcriptional and genome-wide DNA methylation profiling in patients with mild and classic A-T progression as well as healthy controls. RNA and genomic DNA were isolated from peripheral blood mononuclear cells for transcriptional and DNA methylation profiling with RNA-sequencing and modified reduced representation bisulfite sequencing, respectively. We identified 555 genes that were differentially expressed among the control, mild A-T, and classic A-T groups. Genome-wide DNA methylation profiling revealed differential promoter methylation in *cis* with 146 of these differentially expressed genes. Functional enrichment analysis identified significant enrichment in immune, growth, and apoptotic pathways among the methylation-regulated genes. Regardless of clinical phenotype, all A-T participants exhibited downregulation of critical genes involved in B cell function (*PAX5*, *CD79A*, *CD22*, and *FCRL1*) and upregulation of several genes associated with senescence and malignancy, including *SERPINE1*. These findings indicate that gene expression differences may be associated with phenotypic variability and suggest that DNA methylation regulates expression of critical immune response genes in people with A-T.

Ataxia-telangiectasia (A-T) is a rare autosomal recessive disease with a prevalence of approximately 1 in 40,000–100,000 live births<sup>1</sup>. The underlying defect in A-T is absence or reduced production of the ATM protein, a phosphoinositide 3-kinase-like enzyme involved in multiple cellular functions including double stranded DNA repair, cell cycle control, tumor surveillance, telomere length monitoring, and apoptosis<sup>2</sup>. Cells with ATM mutations have higher sensitivity to ionizing radiation that results in increased chromosomal breakage, decreased cell survival<sup>3</sup>, and increased malignancy risk<sup>4</sup>.

Infants with A-T are born phenotypically normal. However, most individuals with A-T develop ataxic symptoms as toddlers, often followed by ocular apraxia, telangiectasias, and loss of ambulatory skills in later childhood<sup>5,6</sup>. Many people with classic A-T also have immune defects. These immunodeficiencies most commonly manifest as low or absent IgA levels, low circulating B and T cell numbers, and suboptimal antibody responses to vaccines<sup>7–9</sup>. Nevertheless, a subset of individuals with A-T exhibit functionally normal immune function with milder and slower neurological progression<sup>10</sup>. Despite this contrast in immune profile, individuals with this mild (atypical) A-T phenotype have a higher risk for cancer, similar to the risk for individuals with classic A-T. Indeed, up to 30% of people with A-T will develop cancer during their lifetime<sup>11</sup>. Lymphoid malignancies are

<sup>1</sup>Eudowood Division of Pediatric Respiratory Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA.

<sup>2</sup>Division of Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>3</sup>A-T Children's Project, Coconut Creek, FL, USA. <sup>4</sup>Eudowood Division of Pediatric, Allergy and Immunology, Johns Hopkins School of Medicine, Baltimore, MD, USA. <sup>5</sup>Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>6</sup>Simpson Querrey Center for Epigenetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ✉e-mail: [smcgrath@jhmi.edu](mailto:smcgrath@jhmi.edu)

Participant	Ataxia (<8 yrs)	Loss of walking (<15 yrs)	Oculomotor ataxia (<15 yrs)	Ocular telangiectasia (<15 yrs)	IgA deficiency	Clinical immunodeficiency	% Criteria	Type of AT based on % criteria	Age (years)
AT002	Yes	Yes	NK	NK	Yes	No	0.75	Classic	26
AT005	Yes	Yes	Yes	Yes	NK	No	0.80	Classic	29
AT006	Yes	Yes	Yes	Yes	No	No	0.67	Classic	12
AT007	Yes	Yes	Yes	Yes	Yes	No	0.83	Classic	14
AT008	Yes	Yes	Yes	Yes	Yes	No	0.83	Classic	12
AT001	Yes	No	NK	NK	No	No	0.25	Mild	33
AT003	Yes	No	Yes	Yes	No	No	0.50	Mild	29
AT004	Yes	No	NK	Yes	NK	No	0.50	Mild	31
AT009WT	—	—	—	—	—	—	—	WT	21
AT010WT	—	—	—	—	—	—	—	WT	26
AT011WT	—	—	—	—	—	—	—	WT	26
AT012WT	—	—	—	—	—	—	—	WT	31

**Table 1.** Phenotypic classification of study participants. NK: not known; WT: wild type.

most frequent during the first two decades of life, while solid tumors are more common in adults. Adults with A-T also exhibit symptoms of premature aging, including cardiovascular disease, metabolic syndrome, and glucose intolerance<sup>12</sup>.

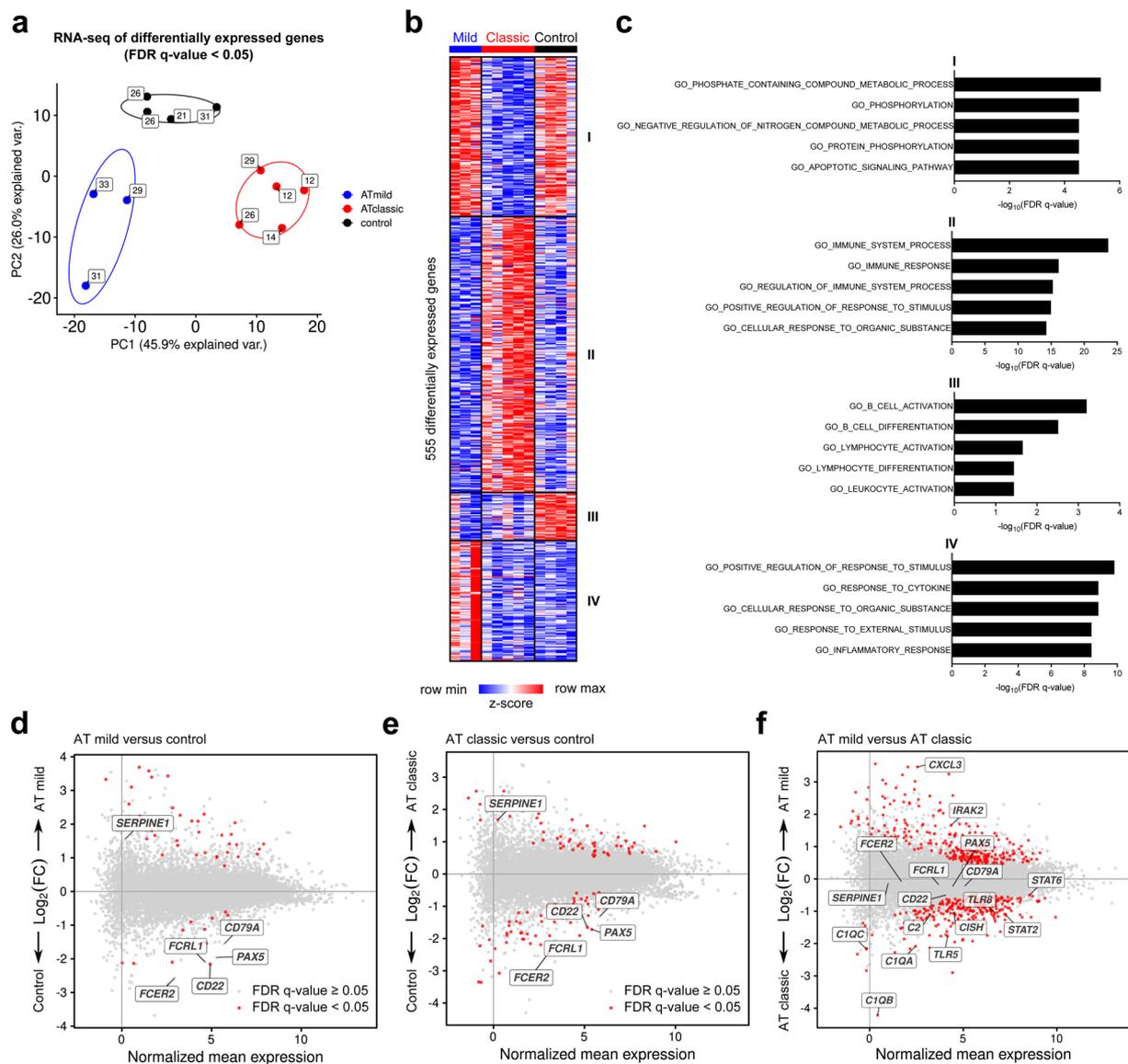
The majority of people with A-T have compound heterozygous mutations of the *ATM* gene, likely contributing in part to the variability in disease presentation and survival<sup>7,12</sup>. Although some genotype-phenotype correlations have been reported in founder populations, mutational hotspots in the *ATM* gene are not universally found<sup>13–16</sup>. In addition, there are no clinical biomarkers or molecular signatures that predict disease progression, malignancy risk, survival, or response to therapy. Having these tools could help with disease surveillance and improve survival for people with A-T.

Identifying molecular pathways associated with specific A-T phenotypes could also be useful in validating biomarkers. This in turn might help identify subsets of people with A-T who may be at higher or lower risk for developing certain disease-specific conditions such as severe infection and malignancy. Identifying genes and gene processes regulated by DNA methylation could also provide insights into regulatory mechanisms that represent therapeutic targets to modulate gene expression and attenuate disease phenotype in people with A-T<sup>17</sup>. Accordingly, we sought to determine the association between phenotypic variability<sup>10</sup> and the transcriptional and epigenetic landscapes in people with A-T. We hypothesized that people with mild (atypical) disease progression and people with classic (more severe) disease progression would have different transcriptional and DNA methylation signatures from each other and from non-A-T healthy control study participants. As there are common phenotypic similarities found among all people with A-T, we also hypothesized that participants with A-T regardless of phenotype would share some common gene signatures. Finally, we hypothesized that genes differentially expressed in people with A-T would be regulated in part by DNA methylation.

## Results

*PBMCs from mild and classic A-T study participants display different transcriptomes from one another and from non-A-T controls.* Eight participants with A-T were assigned to mild or classic phenotypes based on a phenotypic classification described by Fievet *et al.*<sup>10</sup>. Participants characterized as mild A-T had  $\leq 50\%$  of the standard criteria, whereas participants characterized as classic A-T had  $> 50\%$  of the standard criteria ( $p < 0.003$ ) (Table 1 and Supplementary Table S1). Controls included four healthy young adults without A-T. We observed a non-statistically significant variability in age between groups (one-way ANOVA,  $p = 0.051$ ; Sidak's *post hoc* test of non-A-T control group versus classic A-T group,  $p = 0.10$ ). No participants were acutely ill at the time of evaluation. Following grouping of study participants, isolated peripheral blood mononuclear cells (PBMCs) enriched for lymphocytes were used for genome-wide transcriptional and DNA methylation profiling studies with RNA-sequencing (RNA-seq) and modified reduced representation bisulfite sequencing (mRRBS), respectively. We examined complete blood counts obtained from 7/8 A-T participants within less than 1 year of the molecular studies. In participants with A-T, the average lymphocyte percentage was 27.7% (standard deviation, 9.16%; range, 17.0–35.6%) with an average white blood cell count of  $9.4 \times 10^3$  (standard deviation,  $3.18 \times 10^3$ ; range,  $4.08 \times 10^3 - 13.8 \times 10^3$ ). These results are consistent with normal or near-normal peripheral lymphocyte counts. To estimate the relative proportion of B and T cells, we applied an *in silico* cellular deconvolution procedure to the bulk RNA-seq data. This approach revealed decreased frequencies of B and T marker gene profiles among study participants with A-T compared with control participants (Supplementary Fig. S1).

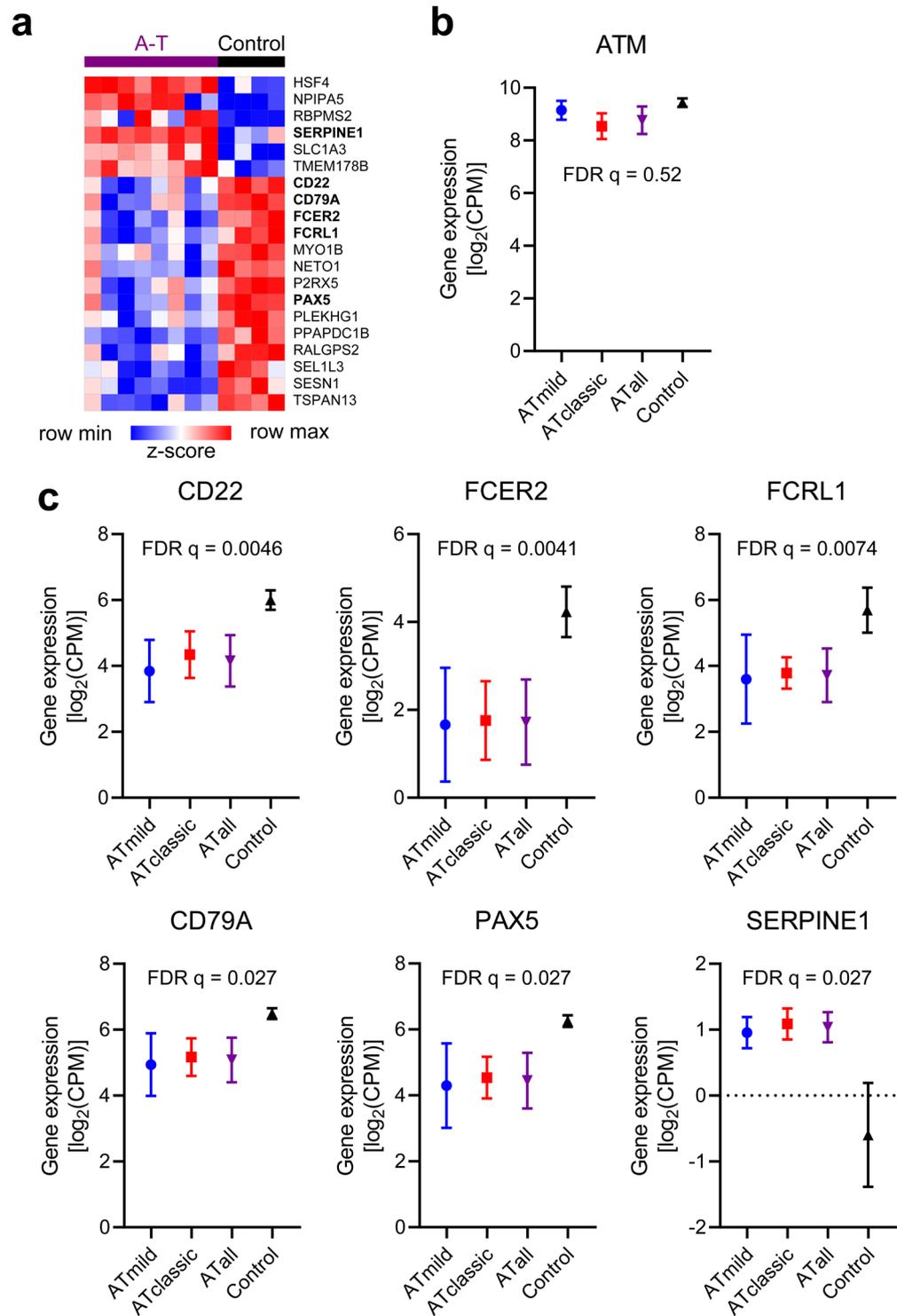
Principal component analysis (PCA) of differentially expressed genes (DEGs) revealed clustering by group assignment (Fig. 1a). PC1 explained the variance between mild and classic A-T phenotypes, and PC2 explained the variance between the A-T and control groups. Multiple group testing with a false-discovery rate (FDR)  $q$ -value  $< 0.05$  revealed 555 DEGs among the three groups, which separated into four  $k$ -means clusters (Fig. 1b and Supplementary Table S2).  $K$ -means cluster 1 contained genes down-regulated primarily among A-T participants with a classic phenotype. Gene ontology (GO) processes in cluster 1 included those involved with apoptosis and regulation of phosphate and nitrogen reactions (Fig. 1c). Gene ontology (GO) processes associated



**Figure 1.** Transcriptional profiling reveals differential gene expression between participants with mild and classic A-T and control study participants. **(a)** Principal component analysis of 555 differentially expressed genes identified from a generalized linear model and ANOVA-like testing with FDR  $q$ -value  $< 0.05$ . Ellipses represent normal contour lines with one standard deviation probability. Points are annotated with participant age in years. **(b)**  $K$ -means clustering of differentially expressed genes with  $k = 4$  and scaled as  $z$ -score across rows. **(c)** Top five gene ontology (GO) processes derived from each  $k$ -means cluster ranked by  $-\log_{10}$ -transformed false-discovery rate (FDR)  $q$ -value. **(d-f)** MA plots comparing the gene expression profile of mild A-T versus control participants (d), classic A-T versus control participants (e), and mild A-T versus classic A-T participants (f). Genes of interest are annotated. FC = fold change.

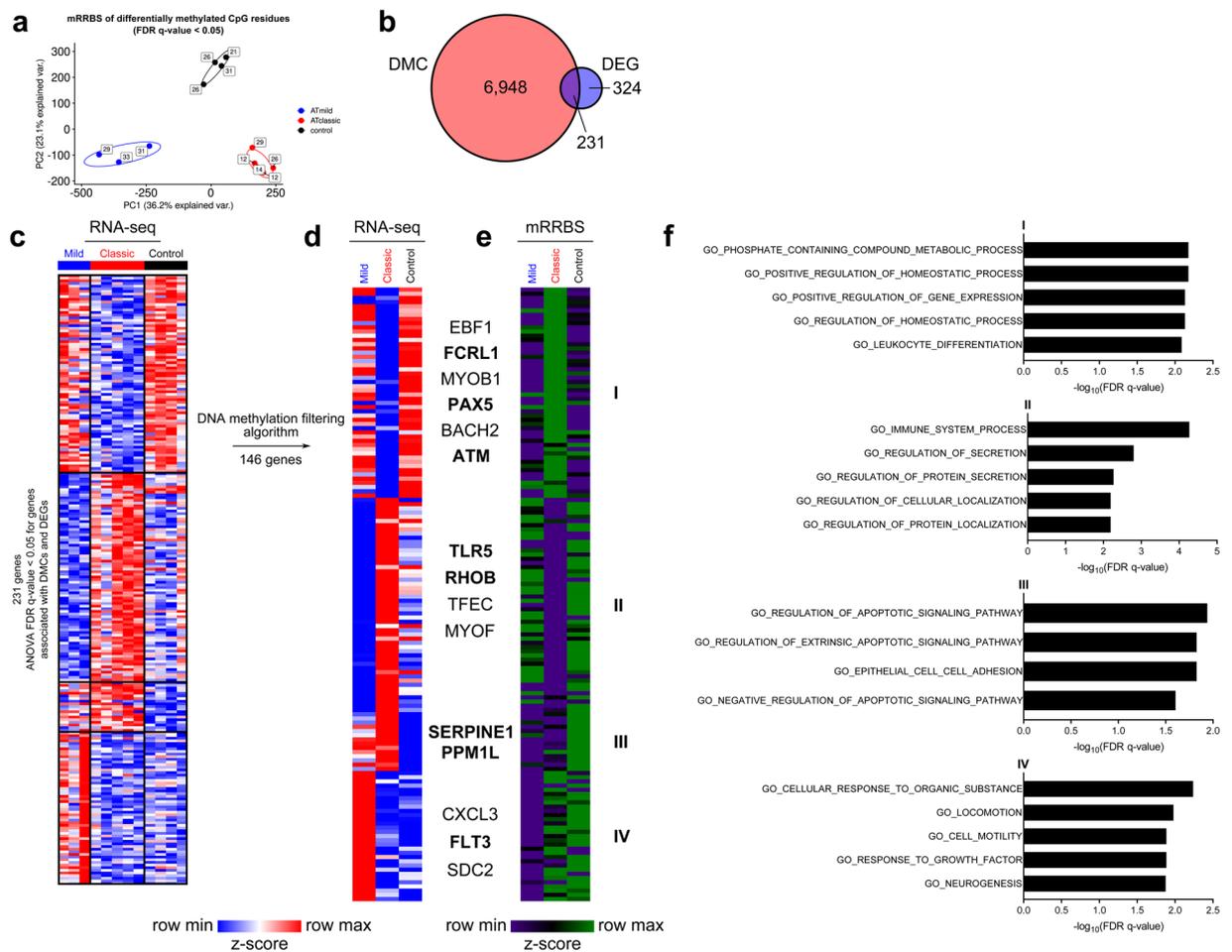
with  $k$ -means cluster 2 were mainly upregulated in the classic A-T participants and included immune system processes.  $K$ -means cluster 3 included GO processes involved with lymphocyte development and B cell morphology and function. These processes were downregulated in both mild and classic A-T participants compared with control participants. In  $k$ -means cluster 4, GO processes involved with rate of response to stimulus and cell activity were primarily upregulated in A-T participants with mild phenotypes. Pairwise comparisons revealed transcriptional differences between A-T groups and control participants as well as between mild and classic A-T phenotypes (Fig. 1d–f).

*A core set of transcriptional responses is common to both mild and classic A-T participants compared with control participants.* We recognized similarities among the mild and classic A-T signatures and identified a core set of 20 DEGs expressed in both mild and classic A-T participants when compared with control participants (Fig. 2a and Supplementary Table S3). *ATM* expression was not significantly different comparing the mild A-T group with non-A-T controls (FDR  $q$ -value  $\approx 0.9$ ); however, the classic A-T group exhibited lower *ATM* expression compared with non-A-T controls (FDR  $q$ -value  $\approx 0.05$ ). Thus, when the A-T groups were merged together, *ATM*



**Figure 2.** A shared transcriptomic signature distinguishes participants with A-T from control participants. **(a)** *K*-means clustering of 20 differentially expressed genes with  $k = 2$  and scaled as z-score across rows. Selected genes are highlighted. **(b)** Gene expression of *ATM* in mild A-T, classic A-T, all A-T, and control participants. **(c)** Gene expression of selected genes in mild A-T, classic A-T, all A-T, and control participants. Plots show mean  $\pm$  standard deviation. FDR  $q$ -values for the pairwise test comparing the merged A-T group with non-A-T controls is shown for **(b)** and **(c)**. CPM = counts per million.

expression was not significantly different from non-A-T controls (Fig. 2b). DEGs downregulated in both mild and classic A-T participants included *PAX5* and *FCRL1* (involved in B cell development), *FCER2* (involved in IgE synthesis), *CD22* (inhibitor of B cell responses), and *CD79A* (involved in transduction of B cell signals) (Fig. 2c)<sup>18–23</sup>.



**Figure 3.** Distinct DNA methylation profiles identify A-T phenotypes. **(a)** Principal component analysis of 66,410 differentially methylated CpGs (DMCs) identified from a beta-binomial regression model with an arcsine link function fitted using the generalized least square method and Wald-test FDR  $q$ -value < 0.05. Ellipses represent normal contour lines with one standard deviation probability. Points are annotated with participant age in years. **(b)** Venn diagram partitioning genes associated with differentially methylated cytosines (DMCs) within 2 kb of their gene bodies (inclusive) and differentially expressed genes (DEGs), all with FDR  $q$ -value < 0.05. **(c)**  $K$ -means clustering of 231 DEGs associated with differentially methylated CpGs (DMCs);  $k = 4$  and scaled as  $z$ -score across rows. **(d)** Results of filtering promoter (transcriptional start site  $\pm 1$  kb) CpGs by 25% difference in CpG methylation between groups. For each  $k$ -means cluster, CpGs with 25% higher methylation in lower expression groups compared with higher expression groups passed the filter. The heat map shows average gene expression ( $\log_2$ -transformed read counts per million) for 146 genes passing the methylation difference filter scaled as  $z$ -score across rows. **(e)** Merged-replicate average CpG methylation in gene promoters scaled as  $z$ -score across rows for the 146 loci passing the methylation difference filter. For (d) and (e), genes are ordered as in (c). **(f)** Top five gene ontology (GO) processes derived from each  $k$ -means cluster ranked by  $-\log_{10}$ -transformed FDR  $q$ -value for the genes in (e and f).

Of additional interest was the finding that *SERPINE1* was upregulated in all A-T participants regardless of phenotype (Figs. 1d,e and 2c). *SERPINE1* encodes plasminogen activator inhibitor-1 (PAI-1), a protein elevated in conditions of premature aging, insulin resistance, and coronary heart disease<sup>24</sup>. Other genes upregulated among A-T participants included *HSF4*, *NPIPA5*, *RBPM52*, *SLC1A3*, and *TMEM178B*. Functions of these genes include regulation of heat shock proteins, stimulation of visceral smooth muscle proliferation, and episodic ataxia with gain-of-function mutations<sup>25–27</sup>. The shared DEGs identified among participants with A-T included genes linked to conditions well described in patients with A-T, including early senescence, the presence of ataxia, the development of malignancy, B cell immune dysregulation, and metabolic syndrome<sup>28</sup>.

DNA methylation is statistically likely to regulate gene expression differences in A-T. To determine whether DNA methylation is associated with differential gene expression among these groups, we performed genome-wide DNA methylation profiling using modified reduced representation bisulfite sequencing (mRRBS)<sup>17,29</sup>. PCA revealed differentially methylated cytosine-phospho-guanine residues (CpGs) between mild A-T, classic A-T, and non-A-T controls (Fig. 3a) that resembled the PCA of DEGs shown in Fig. 1a. We identified 231 genes that were both differentially expressed in the RNA-seq data set and whose promoters were differentially methylated

in the mRRBS data set (Fig. 3b). We then applied a previously-published DNA methylation difference-filtering algorithm designed to identify methylation-regulated genes within this group<sup>29,30</sup>. This procedure identified 146 genes (Fig. 3c–e and Supplementary Table S4) including *FCRL1*, *PAX5*, *SERPINE1*, *PPM1L*, *TLR5*, *RHOB*, and *FLT3*. *K*-means cluster 1 genes were enriched for GO processes involved in homeostatic and regulatory processes (Fig. 3f). GO processes in *k*-means cluster 2 included immune system and regulatory processes involved with protein secretion and localization. *K*-means cluster 3 included processes involved in apoptotic signaling, and *k*-means cluster 4 included processes involved in cell motility and growth. A network analysis of the gene body loci for these 146 genes highlighted immune dysregulation (Supplementary Fig. S2). Collectively, these data sets reveal methylation-regulated gene expression differences both between phenotypes of A-T and between A-T and control participants.

## Discussion

In this study, we found that variability in A-T phenotype correlated with changes in gene expression and epigenetic signature. These findings suggest that differences in transcriptional programs may underlie the phenotypic variability observed in people with mild and classic A-T. We also found a core set of DEGs that were shared by both mild and classic phenotypic participants but differed from non-A-T controls. These shared DEGs included genes involved in B cell development, senescence, and tumor suppression—processes known to be dysregulated among people with A-T<sup>1</sup>. Additionally, we found an association between *cis* promoter site methylation and gene expression in a subset of immune response genes, indicating that epigenetic mechanisms may regulate gene expression of critical immune response genes among people with A-T.

Transcriptional profiling revealed distinct expression signatures between mild and classic A-T phenotypes. This separation between groups accounted for the plurality of variance observed in the principal component analysis and indicated that mild and classic A-T participants were transcriptionally different from each other.

ATM is part of the DNA damage response and maintains genomic integrity by responding to DNA damage through p53-induced cell cycle arrest, repair, and apoptosis. It has previously been shown that apoptosis is activated through ATM via miR-34a/HDAC1<sup>31,32</sup>. ATM is a regulator of DNA-PKcs, which binds to DNA ends and senses breakage<sup>33</sup>. In addition to the ATM pathway, DNA-PKcs can signal through other pathways including c-ABL, p21, and Tap73, which can initiate apoptosis through PUMA, NOXA, and BAX. Thus, apoptosis would be expected to be dysregulated in ATM-deficient individuals, since the overall DNA damage response is altered. In our study, when including all DEGs, participants with classic A-T had lower functional enrichment in apoptotic signaling pathways compared with mild A-T and non-A-T controls. However, when examining only DEGs associated with differentially methylated cytosines (DMCs), we found increased enrichment in apoptotic processes in mild and classic A-T participants compared with non-A-T controls. The participants with classic A-T phenotypes also displayed enrichment of immune processes compared with mild A-T and non-A-T controls. DEGs upregulated in classic A-T participants included the immune response genes *TLR5*, *TLR8*, *STAT2*, and *STAT6* as well as genes encoding complement components (*C1QA*, *C1QB*, *C1QC*, and *C2*). *CISH*, a gene encoding protein that negatively regulates cytokine signaling<sup>34</sup>, was also upregulated among classic A-T participants. These findings suggest that people with a classic A-T phenotype exhibit an upregulated immune response at baseline whereas those with mild A-T phenotypes do not. This dysregulated immune program may be compensatory for the ineffective host response observed among people with classic A-T.

Although ATM protein levels were not measured in this study, *ATM* gene expression was higher in mild A-T participants compared with those with classic A-T. The participants with mild A-T also had higher expression of genes involved in inflammation, including *CXCL3* and *IRAK2*, and GO processes associated with upregulation of cell migration and cell growth. These inflammatory and growth signatures found in participants with mild A-T but not classic A-T could signify an adaptive response to DNA injury unique to people with mild A-T<sup>35,36</sup>. This inflammatory transcriptional signature in the mild A-T participants could also contribute to the higher malignancy risk reported among people with A-T and carriers of *ATM* mutations<sup>37–40</sup>.

We identified a core set of differential transcriptional responses present in both mild and classic A-T participants. *PAX5*, *CD79A*, *CD22*, and *FCRL1* are members of a core set of genes that are critical for B cell development and function<sup>41</sup>. These genes were downregulated among all participants with A-T compared with non-A-T controls. Furthermore, in addition to the role of *PAX5* in B cell development, this gene also acts as a tumor suppressor<sup>22,42</sup>. Indeed, downregulation and dysregulation of *PAX5* may contribute to the higher risk of malignancy, particularly B cell lymphomas, observed among people with A-T<sup>43,44</sup>. We also found that *SERPINE1* was upregulated among all participants with A-T compared with non-A-T controls. *SERPINE1* has been shown to mediate cellular senescence, and a null mutation in *SERPINE1* has been associated with longer telomere length, lower prevalence of diabetes, and longer life span in an extended Amish kindred with a null mutation<sup>24</sup>. Supporting a possible link between elevated *SERPINE1* and A-T phenotype are the observations that people with A-T have accelerated telomere shortening<sup>45</sup>, and that premature aging is increasingly recognized in A-T as overall life expectancies have increased with effective supportive care<sup>12</sup>.

Using a previously published computational tool<sup>29,30</sup>, we found that *cis* DNA methylation was associated with gene expression at the loci of 146 genes that are predominantly involved with immune responses. This finding indicates that DNA methylation differs between participants with mild and classic A-T phenotypes and non-A-T controls and provides insights into the mechanisms that regulate gene expression in A-T. Recognition of DNA methylation as a regulator of gene expression may offer opportunities to modulate abnormal gene expression among people with A-T. These findings also raise mechanistic questions regarding the link between DNA damage and differential methylation patterning that are beyond the scope of our study.

As A-T is a rare disease, limitations to our study include the small number of participants in each group. Nevertheless, the differences in transcriptional signatures between the two A-T groups and non-A-T controls were striking. While some DEGs were similar between A-T phenotypes, many of the DEGs and cellular processes

diverged between the two A-T phenotypes, consistent with the observed variability in clinical phenotypes present in people with A-T. The non-A-T controls in this study were older than the classic A-T participants. Although the difference in age was not statistically significant, age could represent a potential confounder in this study. We used an isolation approach to enrich for lymphocytes. Nevertheless, this approach would not necessarily normalize frequencies of B and T cell subsets between groups; indeed, our deconvolution procedure revealed a decreased proportion of B and T cell marker gene profiles comparing the A-T groups with the control group. These results can be interpreted to mean a decrease in the representation of these cell types in the cells subjected to bulk RNA-seq. Alternatively, as the analysis relies on cell type-specifying marker genes, these results may also represent a decrease in expression of these genes in otherwise balanced cell populations. Regardless, in this study each sample consisted of a composite of lymphocyte populations, representative of the group. Additionally, the use of bulk RNA-seq on mixed cell populations allowed for aggregate measures to reflect the distinct milieus of mild and classic A-T and to include rare populations of cells. Future studies examining molecular signatures in specific lymphocyte subsets may provide a more granular perspective into the role of ATM in immune regulation.

In summary our findings indicate that gene expression differences are associated with phenotypic variability and that DNA methylation may regulate the transcriptional expression of several critical immune response genes among people with A-T.

## Methods

**Participants.** The work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for studies involving humans. The institutional review board of the Johns Hopkins Medical Institutions approved the study (IRB NA\_00051764), and written informed consent was obtained from every participant and/or his/her guardian. All participants with A-T met the diagnostic criteria for A-T based on clinical symptoms and laboratory findings of either elevated alpha-fetoprotein, diminished ATM protein, pathogenic *ATM* mutations, and/or increased chromosomal breakage after *in vitro* exposure to x-rays, as previously established<sup>46</sup>. Demographic information was obtained from chart review. Participants were recruited sequentially from the outpatient Johns Hopkins A-T Clinic and assigned to either a mild or classic A-T phenotype based on a phenotypic classification previously described by Fievet *et al.*<sup>10</sup>. Healthy non-A-T young adults were recruited as controls.

**Isolation of peripheral blood mononuclear cells (PBMCs).** PBMCs were isolated using the Ficoll-Paque Plus Method (GE Healthcare Bio-sciences AB) using a process that maintains the viability of B and T lymphocytes to account for relative lymphopenia among A-T versus control participants as previously described<sup>47</sup>. Briefly, a 7-mL venous blood sample was transferred into a 50-mL conical tube and PBS was added to achieve a 20-mL volume. 15 mL of Ficoll-Paque Plus was transferred into a separate 50-mL conical tube using a syringe. The 20 mL of diluted blood was gently layered onto the Ficoll. The sample was spun at 400 ×g for 30 minutes at room temperature with no brake. The PBMC layer was then collected at the diluted plasma/Ficoll interface. PBS was added to the PBMCs to bring the volume up to 20 mL. The sample was then spun at 200 ×g for 10 minutes at room temperature. The supernatant was discarded and the cells were resuspended in 1 mL of PBS and counted using a TC20 automatic cell counter (Bio-Rad). 4 mL of PBS was added to the sample and spun at 200 ×g for 10 minutes to remove any contaminating Ficoll, platelets, and plasma proteins. For transcriptional and DNA methylation profiling, the sample was resuspended in 350 μL of RLT plus (Qiagen) containing 1% beta-mercaptoethanol, vortexed for 30 seconds, and then transferred to a −80 °C freezer.

**RNA sequencing and analysis.** RNA was isolated using the Qiagen AllPrep DNA/RNA Micro kit. Transcriptional profiling was performed as previously described<sup>30,47–49</sup> with the following modifications. Libraries for RNA sequencing were prepared with the NEBNext Ultra I RNA Library Prep Kit for Illumina with poly(A) mRNA selection (12 ng input) and sequenced using 1 ×75 single-end reads on an Illumina NextSeq 500 instrument using a NextSeq 500/550 V2 High Output reagent kit. Fastq files were trimmed using Trimmomatic v0.38 to remove end nucleotides with a phred score less than 30 while requiring a minimum length of 20 bp and then aligned to the hg38 reference genome using Tophat v2.1.0 with default parameters. Code used for RNA-seq data processing is available from <https://github.com/ebartom/NGSbartom>. The RNA-seq quantitation pipeline in SeqMonk v1.44.0 was used to generate gene and transcript expression levels for protein-coding genes (counts data for uniquely mapped reads over exons). Lowly expressed genes were excluded by including only those genes with more than 1 count per million in at least 3 of the samples. Library sizes were re-computed after filtering these genes, and sample normalization factors were calculated and applied using the trimmed mean of M values (TMM) procedure as implemented in edgeR v3.24.3 (R v3.5.1 with RStudio v1.2.1578). ANOVA-like testing and pairwise comparisons (either treating A-T participants by phenotype [Fig. 1] or merging all A-T participants [Fig. 2], respectively) were performed by fitting a generalized linear model and executing a likelihood ratio test followed by false-discovery rate (FDR) correction using the Benjamini-Hochberg method in edgeR<sup>50</sup>. An FDR *q*-value cutoff of 0.05 was used to identify differentially expressed genes. Functional enrichment analysis for Gene Ontology (GO) biologic processes was performed using Molecular Signatures Database (MSigDB) (The Broad Institute).

Bulk RNA-seq deconvolution was performed using the 10X Genomics PBMC 3K data set as a single-cell RNA-seq reference (<https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k>). The PBMC reference had been previously processed and clustered into seven distinct cell-types using Seurat<sup>51</sup>. Reference counts were library-size corrected and log-transformed with SCANPY<sup>52</sup>, and 2,000 highly variable genes were identified and used for downstream analysis. AutoGeneS selected 600 marker genes through 5,000 generations of its optimizer that minimized correlation and maximized Euclidean distance amongst the

gene-set<sup>53</sup>. These marker genes were then used to estimate cell type proportions within the bulk RNA-seq data set. A Mann-Whitney U test was used to test significance between the merged A-T and control groups.

**Modified reduced representation bisulfite sequencing and analysis.** Genome-wide DNA methylation profiling was performed using modified reduced representation bisulfite sequencing (mRRBS) as previously described<sup>29,30,48,49,54,55</sup>. “Modified” refers to bisulfite conversion of MspI-digested and size-selected DNA prior to random priming, adapter ligation, and indexing; see Singer (2019) for detailed explanation<sup>29</sup>. 250–300 ng of genomic DNA was used as input. Code for mRRBS data processing can be found in the supplement to Singer (2019)<sup>29</sup>. Bisulfite conversion efficiency averaged 99.6% (standard deviation, 0.05%) as estimated by the measured percent of unmethylated CpGs in  $\lambda$ -bacteriophage DNA (New England BioLabs N3013S) added at a 1:200 mass ratio to each sample. In total, we determined the methylation status of 6,697,115 unique CpGs (nearly 25% of the CpG methylome). Beta scores at individual CpG positions were quantified using the SeqMonk platform with the bisulphite feature methylation pipeline. Bismark coverage (counts) files for cytosines in CpG context were analyzed with respect to differential methylation with the DSS v2.34.0 R/Bioconductor package. To detect differentially methylated CpGs, we used DSS to generate a beta-binomial regression model with an arcsine link function fitted using the generalized least square method and then applied a Wald-test with FDR  $q$ -value <0.05. Putative differentially methylated genes were identified as those containing a differentially methylated cytosine in their promoter, defined as the transcriptional start site  $\pm$  1 kb. Our previously described methylation difference-filtering algorithm then selected for promoter CpGs with a methylation directionality (hypermethylated or hypomethylated) inverse to the corresponding gene expression directionality (down- or up-regulated) between groups<sup>29,30</sup>. We used a threshold of 25% for the difference-filtering procedure to establish the list of 146 methylation-regulated genes. Using the gene body coordinates of the 146 methylation-regulated genes, we performed network analysis via the GREAT tool<sup>56</sup> with the basal+extension association rule (constitutive 5 kb upstream and 1 kb downstream, up to 1000 kb max extension) against the whole genome as background.

**Statistical analysis.** Principal component analysis was performed with the `prcomp` base R statistical function.  $K$ -means clustering and heat maps were generated using the Morpheus web interface (<https://software.broadinstitute.org/morpheus/>).  $K$  was selected using the elbow method. Venn diagrams were created with the VennDiagram v1.6.20 R package. All indicated tests were two tailed unless otherwise stated. Computational analysis was performed using “Genomics Nodes” and “Analytics Nodes” on Quest, Northwestern University’s High-Performance Computing Cluster.

### Data availability

The raw processed sequencing data set (counts and coverage tables) is available from the GEO database under accession number GSE142844 (<https://www.ncbi.nlm.nih.gov/geo/>). For other original data, please contact the corresponding author.

Received: 19 February 2020; Accepted: 13 April 2020;

Published online: 04 May 2020

### References

- Rothblum-Oviatt, C. *et al.* Ataxia telangiectasia: A review. *Orphanet J. Rare Dis.* **11**, 159, <https://doi.org/10.1186/s13023-016-0543-7> (2016).
- Mavrou, A., Tsangaris, G. T., Roma, E. & Kolialexi, A. The ATM gene and ataxia telangiectasia. *Anticancer Res.* **28**, 401–405 (2008).
- Stanley, S. E., Rao, A. D., Gable, D. L., McGrath-Morrow, S. & Armanios, M. Radiation sensitivity and radiation necrosis in the short telomere syndromes. *Int. J. Radiat. Oncol. Biol. Phys.* **93**, 1115–1117, <https://doi.org/10.1016/j.ijrobp.2015.08.048> (2015).
- Riaz, I. B., Faridi, W., Patnaik, M. M. & Abraham, R. S. A systematic review on predisposition to lymphoid (B and T cell) neoplasias in patients with primary immunodeficiencies and immune dysregulatory disorders (inborn errors of immunity). *Front. Immunol.* **10**, 777, <https://doi.org/10.3389/fimmu.2019.00777> (2019).
- Devaney, R., Pasalodos, S., Suri, M., Bush, A. & Bhatt, J. M. Ataxia telangiectasia: Presentation and diagnostic delay. *Arch. Dis. Child.* **102**, 328–330, <https://doi.org/10.1136/archdischild-2016-310477> (2017).
- Amirifard, P., Ranjouri, M. R., Yazdani, R., Abolhassani, H. & Aghamohammadi, A. Ataxia-telangiectasia: A review of clinical features and molecular pathology. *Pediatr. Allergy Immunol.* **30**, 277–288, <https://doi.org/10.1111/pai.13020> (2019).
- Pereira, C. T. M. *et al.* B-cell subsets imbalance and reduced expression of cd40 in ataxia-telangiectasia patients. *Allergol. Immunopathol. (Madr.)* **46**, 438–446, <https://doi.org/10.1016/j.aller.2017.09.031> (2018).
- Driessen, G. J. *et al.* Antibody deficiency in patients with ataxia telangiectasia is caused by disturbed B- and T-cell homeostasis and reduced immune repertoire diversity. *J. Allergy Clin. Immunol.* **131**, 1367–1375 e1369, <https://doi.org/10.1016/j.jaci.2013.01.053> (2013).
- Kraus, M. *et al.* Disturbed B and T cell homeostasis and neogenesis in patients with ataxia telangiectasia. *J. Clin. Immunol.* **34**, 561–572, <https://doi.org/10.1007/s10875-014-0044-1> (2014).
- Fievet, A. *et al.* Functional classification of ATM variants in ataxia-telangiectasia patients. *Hum. Mutat.* **40**, 1713–1730, <https://doi.org/10.1002/humu.23778> (2019).
- Choi, M., Kipps, T. & Kurzrock, R. ATM mutations in cancer: Therapeutic implications. *Mol. Cancer Ther.* **15**, 1781–1791, <https://doi.org/10.1158/1535-7163.MCT-15-0945> (2016).
- Shiloh, Y., Lederman, H. M. & Ataxia-telangiectasia, A.-T. An emerging dimension of premature ageing. *Ageing Res Rev* **33**, 76–88, <https://doi.org/10.1016/j.arr.2016.05.002> (2017).
- Suspitsin, E. *et al.* ATM mutation spectrum in russian children with ataxia-telangiectasia. *Eur. J. Med. Genet.*, <https://doi.org/10.1016/j.ejmg.2019.02.003> (2019).
- Chessa, L. *et al.* Founder effects for ATM gene mutations in italian ataxia telangiectasia families. *Ann. Hum. Genet.* **73**, 532–539, <https://doi.org/10.1111/j.1469-1809.2009.00535.x> (2009).
- Laake, K. *et al.* Identical mutation in 55% of the ATM alleles in 11 norwegian at families: Evidence for a founder effect. *Eur. J. Hum. Genet.* **6**, 235–244, <https://doi.org/10.1038/sj.ejhg.5200181> (1998).
- van Os, N. J. H. *et al.* Genotype-phenotype correlations in ataxia telangiectasia patients with ATM c.3576g>a and c.8147t>c mutations. *J. Med. Genet.* **56**, 308–316, <https://doi.org/10.1136/jmedgenet-2018-105635> (2019).

17. Morales-Nebreda, L., McLafferty, F. S. & Singer, B. D. DNA methylation as a transcriptional regulator of the immune system. *Transl Res* **204**, 1–18, <https://doi.org/10.1016/j.trsl.2018.08.001> (2019).
18. Nutt, S. L., Eberhard, D., Horcher, M., Rolink, A. G. & Busslinger, M. Pax5 determines the identity of B cells from the beginning to the end of B-lymphopoiesis. *Int. Rev. Immunol.* **20**, 65–82 (2001).
19. Li, F. J. *et al.* Emerging roles for the *fcr1* family members in lymphocyte biology and disease. *Curr. Top. Microbiol. Immunol.* **382**, 29–50, [https://doi.org/10.1007/978-3-319-07911-0\\_2](https://doi.org/10.1007/978-3-319-07911-0_2) (2014).
20. Rosenwasser, L. J., Meng, J., Chan, M. A., Gigliotti, N. M. & May, B. E. The role of *cd23* in ige dependent signaling: Implications from pharmacogenetics. *Trans. Am. Clin. Climatol. Assoc.* **122**, 27–33 (2011).
21. Tsubata, T. Inhibitory B cell co-receptors and autoimmune diseases. *Immunological medicine*, 1–9, <https://doi.org/10.1080/2578582.6.2019.1660038> (2019).
22. Leblanc, N., Harquail, J., Crapoulet, N., Ouellette, R. J. & Robichaud, G. A. Pax-5 inhibits breast cancer proliferation through *mir-215* up-regulation. *Anticancer Res.* **38**, 5013–5026, <https://doi.org/10.21873/anticancer.12820> (2018).
23. Katikaneni, D. S. & Jin, L. B cell MHC class II signaling: A story of life and death. *Human immunology* **80**, 37–43, <https://doi.org/10.1016/j.humimm.2018.04.013> (2019).
24. Khan, S. S. *et al.* A null mutation in *serpine1* protects against biological aging in humans. *Sci Adv* **3**, ea01617, <https://doi.org/10.1126/sciadv.a01617> (2017).
25. Nakai, A. *et al.* *Hsf4*, a new member of the human heat shock factor family which lacks properties of a transcriptional activator. *Mol. Cell. Biol.* **17**, 469–481, <https://doi.org/10.1128/mcb.17.1.469> (1997).
26. Notarnicola, C. *et al.* The RNA-binding protein *rbpm2* regulates development of gastrointestinal smooth muscle. *Gastroenterology* **143**, 687–697 e689, <https://doi.org/10.1053/j.gastro.2012.05.047> (2012).
27. Iwama, K. *et al.* A novel mutation in *slc1a3* causes episodic ataxia. *J. Hum. Genet.* **63**, 207–211, <https://doi.org/10.1038/s10038-017-0365-z> (2018).
28. Telatar, M. *et al.* Ataxia-telangiectasia: Identification and detection of founder-effect mutations in the ATM gene in ethnic populations. *Am. J. Hum. Genet.* **62**, 86–97, <https://doi.org/10.1086/301673> (1998).
29. Singer, B. D. A practical guide to the measurement and analysis of DNA methylation. *Am. J. Respir. Cell Mol. Biol.* **61**, 417–428, <https://doi.org/10.1165/rcmb.2019-0150TR> (2019).
30. McGrath-Morrow, S. A. *et al.* DNA methylation regulates the neonatal CD4(+) T-cell response to pneumonia in mice. *J. Biol. Chem.* **293**, 11772–11783, <https://doi.org/10.1074/jbc.RA118.003589> (2018).
31. Gupta, S., Silveira, D. A. & Mombach, J. C. M. ATM/*mir-34a-5p* axis regulates a p21-dependent senescence-apoptosis switch in non-small cell lung cancer: A boolean model of *g1/S* checkpoint regulation. *FEBS Lett.*, <https://doi.org/10.1002/1873-3468.13615> (2019).
32. Magni, M., Buscemi, G. & Zannini, L. Cell cycle and apoptosis regulator 2 at the interface between DNA damage response and cell physiology. *Mutat. Res.* **776**, 1–9, <https://doi.org/10.1016/j.mrrev.2018.03.004> (2018).
33. Fielder, E., von Zglinicki, T. & Jurk, D. The DNA damage response in neurons: Die by apoptosis or survive in a senescence-like state? *Journal of Alzheimer's disease: JAD* **60**, S107–s131, <https://doi.org/10.3233/jad-161221> (2017).
34. Palmer, D. C. *et al.* *Cish* actively silences TCR signaling in CD8+ T cells to maintain tumor tolerance. *J. Exp. Med.* **212**, 2095–2113, <https://doi.org/10.1084/jem.20150304> (2015).
35. Gao, B., Ahmad, M. F., Nagy, L. E. & Tsukamoto, H. Inflammatory pathways in alcoholic steatohepatitis. *J. Hepatol.* **70**, 249–259, <https://doi.org/10.1016/j.jhep.2018.10.023> (2019).
36. Hundertmark, J., Krenkel, O. & Tacke, F. Adapted immune responses of myeloid-derived cells in fatty liver disease. *Front. Immunol.* **9**, 2418, <https://doi.org/10.3389/fimmu.2018.02418> (2018).
37. See, A. L., Chong, P. K., Lu, S. Y. & Lim, Y. P. *Cxcl3* is a potential target for breast cancer metastasis. *Curr. Cancer Drug Targets* **14**, 294–309 (2014).
38. Rhyasen, G. W. & Starczynowski, D. T. *Irak* signalling in cancer. *Br. J. Cancer* **112**, 232–237, <https://doi.org/10.1038/bjc.2014.513> (2015).
39. van Os, N. J. *et al.* Health risks for ataxia-telangiectasia mutated heterozygotes: A systematic review, meta-analysis and evidence-based guideline. *Clin. Genet.* **90**, 105–117, <https://doi.org/10.1111/cge.12710> (2016).
40. Jerzak, K. J., Mancuso, T. & Eisen, A. Ataxia-telangiectasia gene (ATM) mutation heterozygosity in breast cancer: A narrative review. *Curr. Oncol.* **25**, e176–e180, <https://doi.org/10.3747/co.25.3707> (2018).
41. Cozma, D. *et al.* B cell activator *pax5* promotes lymphomagenesis through stimulation of B cell receptor signaling. *J. Clin. Invest.* **117**, 2602–2610, <https://doi.org/10.1172/JCI30842> (2007).
42. Chmelarova, M. *et al.* Elevated DNA methylation in malignant tumors of the sinonasal tract and its association with patient survival. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech. Repub.* **162**, 232–238, <https://doi.org/10.5507/bp.2018.012> (2018).
43. Gu, Z. *et al.* Pax5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat. Genet.* **51**, 296–307, <https://doi.org/10.1038/s41588-018-0315-5> (2019).
44. Machida, S. *et al.* Successful treatment of diffuse large B-cell lymphoma in a patient with ataxia telangiectasia using rituximab. *J. Pediatr. Hematol. Oncol.* **35**, 482–485, <https://doi.org/10.1097/MPH.0b013e3182804d59> (2013).
45. Metcalfe, J. A. *et al.* Accelerated telomere shortening in ataxia telangiectasia. *Nat. Genet.* **13**, 350–353, <https://doi.org/10.1038/ng0796-350> (1996).
46. Cabana, M. D., Crawford, T. O., Winkelstein, J. A., Christensen, J. R. & Lederman, H. M. Consequences of the delayed diagnosis of ataxia-telangiectasia. *Pediatrics* **102**, 98–100, <https://doi.org/10.1542/peds.102.1.98> (1998).
47. McGrath-Morrow, S. A. *et al.* Inflammation and transcriptional responses of peripheral blood mononuclear cells in classic ataxia telangiectasia. *PLoS One* **13**, e0209496, <https://doi.org/10.1371/journal.pone.0209496> (2018).
48. Walter, J. M., Helmin, K. A., Abdala-Valencia, H., Wunderink, R. G. & Singer, B. D. Multidimensional assessment of alveolar T cells in critically ill patients. *JCI Insight* **3**, e123287, <https://doi.org/10.1172/jci.insight.123287> (2018).
49. Weinberg, S. E. *et al.* Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature* **565**, 495–499, <https://doi.org/10.1038/s41586-018-0846-z> (2019).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - A practical and powerful approach to multiple testing. *J R STAT Soc B* **57**, 289–300 (1995).
51. Butler, A., Hoffman, P., Smibert, P., Papalexis, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420, <https://doi.org/10.1038/nbt.4096> (2018).
52. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15, <https://doi.org/10.1186/s13059-017-1382-0> (2018).
53. Aliee, H. & Theis, F. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *bioRxiv*, 2020.2002.2021.940650, <https://doi.org/10.1101/2020.02.21.940650> (2020).
54. Wang, L. *et al.* TET2 coactivates gene expression through demethylation of enhancers. *Sci Adv* **4**, eaau6986, <https://doi.org/10.1126/sciadv.aau6986> (2018).
55. Pianti, A. *et al.* CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci Adv* **5**, eaax2887, <https://doi.org/10.1126/sciadv.aax2887> (2019).
56. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501, <https://doi.org/10.1038/nbt.1630> (2010).

## Acknowledgements

SAM-M was supported by The A-T Children's Project and NIH/NHLBI grant R01HL114800. PAR was supported by NIH/NHLBI K08HL146943, the Parker B. Francis Fellowship, and the ATS Foundation/Boehringer Ingelheim Pharmaceuticals Inc. Research Fellowship in IPF. BDS was supported by NIH/NHLBI grants K08HL128867 and R01HL149883. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, Francis Family Foundation, or the American Thoracic Society. This research was supported in part through the computational resources and staff contributions provided by the Genomics Compute Cluster, which is jointly supported by the Feinberg School of Medicine, the Center for Genetic Medicine, and Feinberg's Department of Biochemistry and Molecular Genetics, the Office of the Provost, the Office for Research, and Northwestern Information Technology. The Genomics Compute Cluster is part of Quest, Northwestern University's high performance computing facility, with the purpose to advance research in genomics.

## Author contributions

S.A.M.-M., C.R.-O., J.M.C., J.W., H.M.L. and B.D.S. contributed to the conception, hypotheses delineation, and design of the study; S.A.M.-M., R.N., K.A.H., B.K., P.A.R. and B.D.S. performed experiments/data acquisition or analysis; S.A.M.-M., H.M.L. and B.D.S. wrote the manuscript or provided substantial involvement in its revision.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-64514-2>.

**Correspondence** and requests for materials should be addressed to S.A.M.-M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020