Glucocorticoid-regulated bidirectional enhancer RNA transcription pinpoints functional genetic variants linked to asthma

Sarah K Sasse[1*], Amber Dahlin[2*], Lynn Sanford[3], Margaret A. Gruca[3], Arnav Gupta[1,4], Ann Chen Wu[5], Carlos Iribarren[6], Robin D. Dowell[3,7,8], Scott T. Weiss[2], and Anthony N. Gerber[1,4,9]

[*]Equal contribution

[1]Department of Medicine, National Jewish Health, Denver, CO, 80206, USA
[2]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[3]BioFrontiers Institute, University of Colorado, Boulder, CO, 80309, USA.
[4]Department of Medicine, University of Colorado, Aurora, CO, 80045, USA
[5]PRecisiOn Medicine Translational Research (PROMoTeR) Center, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA
[6]Kaiser Permanente Division of Research, Kaiser Permanente, Oakland, CA
[7]Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA.
[8]Computer Science, University of Colorado, Boulder, CO, 80309, USA.
[9]Department of Immunology and Genomic Medicine, National Jewish Health, Denver, CO, 80206, USA.

Corresponding Author
Dr. Anthony N. Gerber
Department of Medicine, National Jewish Health,
Room K729, 1400 Jackson St, Denver, CO,
80206, USA, Tel.: 303-270-2783,
E-mail: gerbera@njhealth.org

Funding sources: NHLBI R01 HL109557 (to A.G.); NHLBI R01 HL152244 (to A.D.); NIH R01 GM125871 (to R.D.)

Keywords: asthma, SNP, glucocorticoid, enhancer, transcription

Conflict of interest disclosure statement: ANG and SKS own shares in Psammiad Therapeutics; RDD is a cofounder of Arpeggio Biosciences; STW receives royalties from UpToDate and is on the Scientific Board of Histolix.

Abstract

Genome wide association studies of asthma have not explained environmental risk or variable clinical efficacy of glucocorticoids. Bidirectional enhancer RNA (eRNA) transcription is a widespread response to environmental signals and glucocorticoids. Therefore, we investigated whether single nucleotide polymorphisms (SNPs) within dynamically regulated enhancer RNA (eRNA)-transcribing regions contribute to genetic variation in asthma. Through applying machine-learning methods to a large clinical cohort, we identified novel associations between asthma and 36 SNPs located in eRNA-transcribing regions implicated in regulating diverse cellular processes relevant to asthma. Functional validation established that *rs258760* (mean allele frequency = 0.35, asthma odds ratio 0.95; $P<0.005$) eliminates an active aryl hydrocarbon receptor (AHR) response element linked to transcriptional regulation of the glucocorticoid receptor by AHR ligands commonly found in air pollution. Our findings establish eRNA signatures as a tool for discovery of functional genetic variants and define a novel link between air pollution, glucocorticoid signaling and asthma.

Introduction

Asthma is a highly prevalent disease for which pathogenesis and treatment responsiveness are governed by complex stochastic interactions between genetic background and the environment[1]. Genome wide association studies (GWAS) have defined numerous single nucleotide polymorphisms (SNPs) and associated pathways linked to risk of developing asthma[2-4]. However, the sum of heritabilities explained by these SNPs is estimated to account for only 2.5% of the estimated genetic risk of asthma[5-7]. Moreover, the majority of asthma-associated SNPs defined though GWAS are in non-protein coding portions of the genome[8], and direct regulatory effects of most such SNPs have not been established[9]. These relative weaknesses of GWAS-based approaches are likely due to several underlying factors, including unaccounted effects of environment on asthma risk, linkage disequilibrium, and the very stringent p-values needed to identify significant genetic associations on a genome-wide basis[10]. These drawbacks have also limited the discovery of genetic variants relevant to specific asthma phenotypes, such as responsiveness to therapeutics. For example, clinical heterogeneity in responses to glucocorticoids in asthma is well described, however, relatively few genetic variants have been discovered that directly modify transcriptional responses to glucocorticoids or that directly link glucocorticoid targets to asthma risk[11].

Specialized regions within the genome that control gene expression, frequently referred to as cis-regulatory elements or enhancers, are enriched for functional genetic variants that are linked to disease heritability[12,13]. As a method to reduce the p-value burden of traditional GWAS and increase the likelihood of identifying putative functional variants, strategies that incorporate enhancer features into genetic association studies have been developed and applied to various diseases[14]. However, enhancers, as defined using standard assays for chromatin accessibility or histone modifications, can span hundreds of kilobases[15], which impacts resolution and statistical power. Moreover, these approaches often fail to define relationships between enhancers and regulation of specific genes[16], especially in the context of disease-associated signaling pathways that reshape both enhancer activity and gene transcription.

Sequencing of non-polyadenylated or nascent transcripts has revealed that RNA is transcribed from active enhancers[17,18]. Enhancer RNA (eRNA) transcription has emerged as a very sensitive marker of enhancer activity[19], and concomitant with regulating gene expression, glucocorticoids and inflammatory signals cause significant changes in eRNA transcription[20,21]. Discrete sites of RNA polymerase II (RNAPII) loading and transcription initiation within enhancers, which frequently serve as a nidus for the dynamic binding of regulatory transcription factors[18,22], including the glucocorticoid receptor, can be bioinformatically inferred based on bidirectional transcription signatures[22]. We hypothesized that regions localized to sites of RNAPII loading within eRNA-transcript signatures that change dynamically in response to anti-inflammatory glucocorticoids or pro-inflammatory signals in airway epithelial cells harbor functional genetic variants associated with asthma risk. Here, we apply machine-learning methods to filter SNPs within a large asthma

cohort based on dynamic eRNA-transcript signatures, and we perform fine mapping, annotation and functional validation of a subset of these novel asthma-associated SNPs.

Results

Localization of SNPs within dynamically regulated enhancers and asthma associations in a large clinical cohort

Through analyzing nascent RNA transcription on a genome-wide basis in BEAS-2B cells, we previously identified a set of 1714 enhancer RNAs (eRNAs) dynamically regulated by the prototypical anti-inflammatory glucocorticoid, dexamethasone (dex), and/or TNF, a canonical inducer of NFkB signaling widely used to model therapeutic glucocorticoid cross-talk with inflammatory signals[20]. We reasoned that regions surrounding sites of RNAPII loading within these dynamically transcribed eRNAs would be enriched for functional genetic variants. Therefore, using Tfit modeling[22], we systematically inferred and annotated the sites of RNAPII loading and bidirectional transcription initiation, frequently referred to as Mu (μ), within this eRNA set. Next, to discern potential relationships between these eRNA-transcribing genomic regions and asthma risk in the Genetic Epidemiology Research in Adult Health and Aging (GERA) clinical population[23,24], we leveraged machine learning and fine mapping approaches, which have advantages over traditional statistical inference methods[25], to create an analysis pipeline (Fig 1A) to discover genetic variants associated with asthma based on proximity to μ. We first screened the GERA dataset to filter SNPs that were annotated within the dynamically regulated enhancer regions, by treatment (dex or TNF). A total of 143,869 SNPs were mapped to dex-responsive enhancer regions and 137,339 were mapped to TNF-responsive enhancer regions. Next, we filtered these enhancer-SNP sets to include only those that were annotated to regions within 50 bp (+/-) of the Tfit-predicted RNAPII loading site, μ (μ-localized SNPs, or μ-SNPs). Within dex-responsive regions, we identified 866 total annotated μ-SNPs (661 upregulated; 205 downregulated), and within TNF-responsive regions we found 826 μ-SNPs (743 upregulated; 83 downregulated). The lower number of μ-SNPs observed within downregulated regions follows the predicted biological function of enhancers, i.e. dex or TNF perturbation are increasingly recognized as directly affecting gene transcription through upregulating enhancer site function[26,27].

The subsets of μ-SNPs grouped by changes in eRNA transcription with treatment were then tested for association with asthma status using multivariate regression models, and the significance of associations was determined using a permutation-based approach implemented in the statistical genetic analysis software, PLINK (v. 1.9)[28]. Across the four treatment groups, 75 μ-SNPs were identified as significantly associated with asthma status (25, dex upregulated; 15, dex downregulated; 32, TNF upregulated; 3, TNF downregulated). As regions in close proximity to μ are also enriched for functional transcription factor binding sites[22], SNPs within these regions were further characterized using *in silico* fine mapping to investigate possible functional impacts.

Ultimately, out of the total set of 1692 μ-SNPs, this analysis pipeline output 36 significantly associated μ-SNPs with compelling evidence for functional roles based on robustness of association with asthma (Table 1) and proximity to μ (Supp Table 1).

Refined mapping and analysis of μ-SNP associations

Confidence in predicting RNAPII loading sites based on Tfit-annotated enhancers depends on the complexity and magnitude of the bidirectional eRNA signature. To focus on SNPs that co-localize with high confidence sites of RNAPII loading (μ), we visually inspected the GRO-seq tracks within each enhancer region of interest to manually annotate and refine the bidirectional calls (Supp File 1). One μ-SNP was excluded here due to lack of visual confirmation of a clear bidirectional signature in its surrounding region. The remaining 35 regions were passed to a custom script that refines the estimate of μ based on each manually refined bidirectional annotation. A total of 24 μ-SNPs (69%) located within 500 bp (+/-) of the refined μ sites were classified as high confidence μ-SNPs (Supp Table 2) and prioritized for further analysis. Notably, of these high confidence μ-SNPs, 20 (83%) resided in enhancers that, based on similar transcriptional patterns ($p_{adj} < 0.05$) in response to dex and/or TNF, could be linked to regulation of at least one gene within 1 Mb (+/-) of the eRNA (Supp Table 2). Gene products linked to individual SNPs in this manner include signaling molecules, such as BMP1 and PLK2, ubiquitin ligases, including HERC4, and several transcription factors (e.g. CEBPB). Examples of high confidence vs lower confidence sites of μ-SNP co-localization are depicted in Figs 1B-C, respectively. Genes linked to individual SNPs

To further assess the regulatory function of the high confidence μ-SNP regions, we cloned ~450-1100 bp spanning 9 of the high confidence μ-SNP regions into a minimal promoter luciferase reporter (Supp Table 3). We assayed responses of these reporters to dex and TNF in BEAS-2B cells in comparison to two control glucocorticoid-responsive reporters and two reporters derived from low confidence μ-SNP regions. We found that 7 of the 9 high confidence enhancer reporters (78%) exhibited dynamic changes in activity (Fig 2A) consistent with regulation of the corresponding endogenous eRNAs (Fig 2B; representative examples). These data indicate that the high confidence μ-SNPs are likely to confer biologically relevant regulatory function.

Functional analysis characterizes novel SNPs relevant to GR signaling

To identify high confidence μ-SNPs that are most likely to exert a functional role in controlling gene expression, we used MatInspector[29] to annotate canonical transcription factor binding motifs within 30 bp (+/-) of each SNP site. Through comparing motifs identified for both the canonical and variant alleles, *rs149411423* and *rs258760* emerged as strong candidates for functional regulation. *rs149411423* is a C>T transversion (MAF in GERA < 0.01; Table 1) located in an enhancer region ~175 kb from the transcription start site (TSS) of *GCNT1*, which encodes a glycosyltransferase implicated in pulmonary immune responses and allergy[30,31]. As

shown in Figure 3A, both *GCNT1* (log_2FC = 2.19, p_{adj} = 3.93E-03) and the enhancer harboring *rs149411423* (log_2FC = 2.57, p_{adj} = 0.003) show increased transcription after dex treatment in BEAS-2B cells, and chromatin accessibility was observed at this region (Fig 3B, *top*) in differentiated human primary airway epithelial cells within ATAC-seq data we generated previously[32], supporting physiological relevance of the region. Further (Fig 3B, *bottom*), dex-inducible GR occupancy was observed at the *rs149411423*-μ-SNP region in our BEAS-2B cell datasets[20,33], and in our ChIP-seq analysis of cultured primary human airway smooth muscle (HASM)[34], another critical effector cell type in asthma. MatInspector analysis of this region indicated that the *rs149411423* transversion site resides within a putative glucocorticoid response element (Figure 3C), and through applying a customized position weight matrix we created to predict relative GR binding affinities[35], we found that the predicted affinity of the sequence containing the major C allele for binding GR is ~2-fold greater than that containing the variant T allele. To directly determine whether *rs149411423* alters GR-regulated transcription, we generated a reporter harboring the variant allele using site-directed mutagenesis and assayed its activity compared to the parent construct with the major C allele in BEAS-2B and primary HASM cells. Relative to the parent construct, the mutant reporter exhibited dramatically attenuated induction by dex in all tested cell types (Fig 3D). In aggregate, these data implicate *rs149411423* as reducing GR-mediated induction of an enhancer that regulates transcription of *GCNT1,* strongly supporting a functional role for this SNP.

The second strong candidate SNP, *rs258760,* is a C>T transversion (MAF in GERA 0.34; Table 1) within an *ARHGAP26* intron. This genomic location is approximately 200 kb downstream of the TSS for *NR3C1*, which encodes the glucocorticoid receptor. To assess whether the enhancer region harboring *rs258760* is associated with transcriptional control of *ARHGAP26* or *NR3C1*, we compared eRNA transcription patterns to nascent transcription of both genes after dex or TNF+dex treatment. Fig 4A (GRO-seq tracks) illustrates that eRNA transcription from the *rs258760* region is reduced by dex (log_2FC = -1.8, p_{adj} = 0.035) and TNF+dex treatment (log_2FC = -2.01, p_{adj} = 0.011). Whereas *ARHGAP26* transcription is increased by dex (log_2FC = 0.76, p_{adj} = 0.035) and TNF+dex (log_2FC = 1.41, p_{adj} = 1.01E-07), *NR3C1* transcription is reduced by treatment with dex (log_2FC = -0.56, p_{adj} > 0.05) and TNF+dex (log_2FC = -1.36, p_{adj} = 0.0008), suggesting that *rs258760* resides in a regulatory element for *NR3C1*. Supporting this notion, in our published analysis of nascent transcriptional responses to wood smoke particles[28], eRNA transcription from the same +200 kb enhancer was increased after 30 minutes of wood smoke particle exposure in BEAS-2B cells (log_2FC = 1.99, p_{adj} = 6.47E-62), with a congruent increase in transcription of *NR3C1* (log_2FC = 0.90, p_{adj} = 9.66E-43) but not *ARHGAP26* (log_2FC = -0.11, p_{adj} > 0.05; Fig 4A, PRO-seq tracks). Together, these data strongly suggest that the enhancer harboring *rs258760* controls transcription of *NR3C1*, rather than *ARHGAP26*. This μ-SNP region also aligns with an ATAC-seq peak in patient-derived airway epithelial cells[32], providing additional support for the physiologic relevance of this enhancer (Fig 4B). Furthermore, our MatInspector analysis of transcription factor binding motifs in this region in the context of the major C vs variant T alleles revealed that

the C allele is a critical residue within a canonical binding motif for the aryl hydrocarbon receptor (AHR). This AHR motif is lost in the context of the T allele (Fig 4C), raising the possibility that AHR ligands, which are found at significant concentrations in wood and wildfire smoke[36-38], control the activity of this enhancer in an allele-specific manner.

To determine whether AHR and *rs258760* regulate activity of this presumptive NR3C1 enhancer, we used site-directed mutagenesis to create a reporter with the minor, asthma-protective T allele, and assayed its activity compared to the parent construct in BEAS-2B cells. As illustrated in Fig 4D, while the parent C allele reporter shows robust induction following exposure to the dioxin, TCDD, a potent AHR ligand responsible for toxic effects of Agent Orange[39], this induction was markedly reduced in the mutant reporter harboring the minor T allele. Thus, *rs258760* eliminates a functional AHR response element linked to regulating expression of *NR3C1* in airway epithelial cells in response to AHR ligands. Furthermore, pre-treatment of BEAS-2B cells with TCDD prior to dex exposure potentiates induction of canonical GR transcriptional targets (Fig 4E), such as *FKBP5*[40]. These findings implicate the major C allele, which is associated with increased risk of asthma, as integral to an active AHR response element controlling a functional AHR-NR3C1 signaling axis in airway epithelial cells.

Discussion

Nascent transcript sequencing is a powerful method for quantifying enhancer activity and has been applied previously to study eQTLs[41,42], yet had not been systematically leveraged to identify novel SNP-disease associations. Our results indicate that constraining genetic association analysis based on SNP proximity to bidirectional eRNA transcription initiation sites can facilitate discovery of genetic variants that influence asthma risk. SNP filtering and analysis based on nascent eRNA transcription signatures offers other unique advantages including enrichment for functional regions within enhancers, which can span thousands of base pairs when defined using other methods, and facilitating the identification of presumptive target genes for specific SNP-enhancer regions. Moreover, although we focused here on SNPs within eRNA-centered RNAPII loading regions that respond dynamically to glucocorticoids or TNF, this approach is generalizable to eRNAs regulated by other stimuli relevant to asthma, such as Type II cytokines or particulate pollution. Thus, our data provide a systematic method for discovery of functional variants that influence asthma or other airway disease risk in the context of specific endotypes or environmental exposures.

Two variants, *rs149411423* and *rs258760,* directly illustrate the power of our approach and extend our understanding of GR signaling in the context of asthma risk. *rs149411423,* a high risk, low frequency allele, disrupts a functional GRE that regulates *GCNT1*, a glycosyltransferase linked to IgE levels, immune signaling, and the pulmonary immune response to infection[30,31,43]. Thus, although it remains to be determined whether

rs149411423 directly influences the therapeutic effects of GCs in asthma, we elucidate a novel and physiologically plausible GR-GCNT1 axis that is relevant to asthma risk.

We also discovered a genetic variant, *rs258760*, a low risk, high frequency allele that regulates expression of the glucocorticoid receptor itself in relationship to asthma risk. The more common C allele, associated with increased risk of an asthma diagnosis, is a required nucleotide within a functional binding site for AHR. Our data implicate this AHR response element in controlling increased transcription of airway epithelial GR in response to AHR ligands, which are present at physiologically relevant concentrations in particulate matter air pollution. Although additional genetic and functional validation would be worthwhile, our data support a model in which increased GR expression and local inflammatory repression in the setting of a functional AHR response element is deleterious in the context of repetitive injurious exposures, thus promoting asthma. Our results also provide a critical genetic link between pollution exposure and development of asthma, a problem of increasing importance given the escalating deleterious effects of climate change on air quality[44-46]. In aggregate, we define novel genetic links between GR signaling and asthma risk while establishing a powerful new application of eRNA transcription signatures for discovery of SNP-disease associations.

METHODS

Nascent transcription and cistromic datasets

The BEAS-2B GRO-seq data initially used to identify differentially transcribed eRNAs and their inferred sites of RNAPII loading (μ) were previously described[20], with linked Gene Expression Omnibus (GEO) accession number GSE124916 (30 min datasets). The GEO accession number for the primary human airway epithelial cell ATAC-seq data (Norm-1 and Norm-2 datasets), described previously[32], is GSE157606. BEAS-2B ChIP-seq data were published previously[33], with GEO accession number GSE79803. HASM ChIP-seq data were also described previously[34], with GEO accession number GSE95632. Here, to create the screenshots depicted in Figure 3, we remapped the BEAS-2B and HASM ChIP-seq datasets to hg38 using our standardized ChIP-seq pipeline. Briefly, reads are trimmed for adapters, minimum length, and minimum quality using the bbduk tool from the BBMap Suite (v. 38.05). Trimmed reads are then mapped to hg38 using hisat2 (v. 2.1.0). Resulting SAM files are converted to sorted BAM files using samtools (v. 1.9) and to bedGraph coverage format using genomeCoverageBed from the BEDTools suite (v. 2.29.2). Read coverage is normalized to reads per million mapped using a custom python script and files converted to TDF format using igvtools (v. 2.5.3) for visualization in IGV. The BEAS-2B 30 minute wood smoke particle exposure PRO-seq data depicted in Figure 4 were published previously[47], with GEO accession number GSE167371.

Genomic data

We utilized genotype (SNP) and phenotype (asthma diagnosis, age, sex, BMI, and other demographic data) information from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, which

includes over 110,000 subjects (68,623 asthma cases) with extensive EMR information and genome-wide SNP genotypes (>8 million typed and imputed markers). As GWAS-based associations of genetic risk factors with disease susceptibility can vary significantly by ethnic group within large multi-ethnic cohorts such as GERA[24], we limited the current study to a subset of the cohort with the highest percentage of asthma cases (54,543 Non-Hispanic Whites; 29.8% asthma cases). A case-control study design was used to determine the association of patient genotype (continuous predictor variable) with asthma status (dichotomized outcome variable).

μ-SNP co-localization and statistical association with asthma risk

Unless otherwise indicated, R statistical computing software and PLINK (v. 1.9)[28] (http://pngu.mgh.harvard.edu/purcell/plink) were used for all analyses.

Machine learning methods.

To identify genetic predictors for asthma that were also present in dynamically regulated enhancer regions, we first screened the GERA subset (54,543 samples) to filter SNPs that were annotated within the Tfit-annotated enhancer regions, by treatment (dex or TNF). Next, we filtered these SNP sets to include only those that were annotated to regions within 50 bp (+/-) of the Tfit-predicted RNAPII loading site (μ), referred to as μ-SNPs. After sorting by minor allele frequency (MAF), μ-SNPs were categorized as common (MAF $\geq$ 0.05) or rare (MAF < 0.05) variants and tested independently. Tests for association of additive genotype (categorical variable) with asthma status (dichotomized outcome) were conducted using multivariate logistic regression models adjusted for covariates (age, sex, BMI, smoking status (never, current, former), and six principal components). The significance of associations was determined using a permutation-based approach. One method to improve feature prediction is to combine model-based approaches with permutation. Since each set of dynamically regulated enhancer regions contained relatively few (<1000) SNPs genotyped in GERA, yielding a small number of tests, far fewer SNPs were interrogated overall as compared to a GWAS, which tests millions of SNPs. In this instance, multiple correction methods traditionally applied in GWAS (e.g. Bonferroni) are overly conservative, or over-correct for Type I error. In contrast to the Chi-square test, permutation-based approaches are appropriate for analysis of rare alleles or smaller numbers of samples and preserve the correlational structure between SNPs, requiring a less stringent multiple test correction threshold. Max(T) permutation was applied with 5000 permutations to generate two sets of empirical (pointwise and familywise) P values. The max(T) permutation approach is advantageous for candidate SNP-based approaches, as it is typically sufficient to perform a much smaller number of tests; therefore, a P value threshold of 0.05 was specified. Following permutation, associations were prioritized based on their level of significance.

Fine Mapping and Candidate SNP Selection. An *in silico* fine mapping approach was applied using a web-based annotation tool, SNPnexus, to determine the SNP genomic localization, predicted functions (coding, non-coding, etc.), predicted consequences on protein/gene expression or function, relationship to regulatory

elements (e.g. transcription factor motifs, CpG windows), and impact on structural variation (e.g. CNVs). Prior reports of associations with asthma, lung function measures, GC response or related phenotypes were also determined through literature search using the NCBI PubMed database. After fine mapping, SNPs were grouped according to whether their respective enhancer regions were up- or down-regulated and compared across groups. The resulting set of SNPs with the strongest predictive evidence for causality and closest proximity to μ were prioritized for functional characterization *in vitro*.

Manual refinement of μ and classification of high confidence μ -SNPs and associated target genes

For each of the 36 candidate μ-SNP's, we visually examined the GRO-seq tracks within the surrounding genomic region in order to manually refine the bidirectional calls originally made by Tfit. We passed these refined regions (Supp File 1) to a custom script that calculates the midpoint between the peak signal intensity in each half of the region, thereby refining the estimate of μ for each bidirectional. We then averaged the μ values for any sets of replicates in which the specified bidirectional had robust signal and compared the refined μ estimate to the SNP location. SNPs located within 500 bp (+/-) of the refined μ locations were classified as high confidence μ-SNPs and examined further (Supp Table 2). A custom python script, in conjunction with bedtools (v. 2.28.0) was used to identify genes within 1 Mb (+/-) of each SNP and append DESeq2 results for easy comparison with the relevant μ-SNP bidirectional signature. To account for the short 30 minute GRO-seq transcription timepoint potentially missing true signal changes in longer gene transcripts, a second DEseq2 analysis was performed. GTF files were generated corresponding to the first 90 kb following the TSS. All genes smaller than 90 kb were retained in full. These values assume a polymerase processivity of 3 kb/min. Reads were then counted across these GTF files using the featureCounts algorithm within the Rsubread R package (v. 3.10, R v. 3.6.1) and analyzed with DESeq2 (v. 1.26.0) for differential expression. Data for *NR3C1*, which spans ~150 kb, are based on this truncated DESeq2 analysis.

Cell culture and Reagents

BEAS-2B transformed normal human airway epithelial cells (ATCC) were grown in DMEM with L-glutamine and 4.5 g/L glucose (Corning) containing 10% FBS (VWR) and 1% penicillin/streptomycin (pen/strep; Corning). Primary human airway smooth muscle (HASM) cells (provided by Dr. Reynold Panettieri, Rutgers Biomedical Health Sciences) were cultured in Ham's F12 Medium with L-glutamine (Corning) containing 10% FBS and 1% pen/strep. All cells were maintained in 5% CO_2 at 37°C.

Dexamethasone (dex; Sigma-Aldrich) was dissolved in sterile 100% ethanol (vehicle) and used at a final concentration of 100 nM. Recombinant Human TNF-alpha Protein (TNF) purchased from R&D Systems was diluted in sterile 1X Dulbecco's phosphate buffered saline (DPBS) containing 0.1% bovine serum albumin

(BSA) and used at a final concentration of 20 ng/ml. 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) obtained from Cambridge Isotope Laboratories was diluted in sterile DMSO (vehicle) and used at a final concentration of 10 nM.

Cloning, transfection and luciferase assays

Enhancer reporter constructs were amplified from BEAS-2B genomic DNA by PCR, cloned into pCR2.1-TOPO (Life Technologies), and subsequently ligated into the pGL3-Promoter vector backbone (Life Technologies) using SacI/XhoI (p*COL8A1*) or KpnI/XhoI (all others). p*FKBP5*, p*TNFAIP3*, and p*MT2A* positive control constructs have been described[35,40,48]. The QuikChange II Site-directed Mutagenesis (SDM) Kit from Agilent Technologies was used as instructed by the manufacturer to generate reporter constructs harboring variant alleles (p*GCNT1* C>T and p*NR3C1* C>T). PCR primer sequences used for cloning and SDM and detailed information pertaining to each reporter construct and related μ-SNP are presented in Supp Table 3.

For transfection experiments, BEAS-2B cells were seeded on 96-well tissue culture plates at 40K cells/well in complete medium. The next day, cells were transfected with a total of 200 ng DNA/well (10:1 ratio of firefly luciferase reporter construct to *Renilla* luciferase internal control (pSV40-RL; Promega)) using Lipofectamine 2000 (0.5 ul/well) from Life Technologies. Transfection complexes were removed the following day and cells were treated in fresh complete medium for 8 hr prior to luciferase assay. HASM1 cells were seeded on 96-well tissue culture plates at 50K cells/well in complete medium. Medium was replaced two days later with fresh complete and the following day, cells were transfected with a total of 125 ng DNA/well (10:1:1 ratio of firefly luciferase to renilla luciferase to mCherry plasmid, respectively) using Lipofectamine 3000 (0.3 ul/well) and P3000 Reagent (0.8 ul/well) from Life Technologies. Transfection complexes were removed the following day and cells were treated in fresh complete medium for 24 hr prior to luciferase assay. HASM2 cells were seeded on 96-well plates at 40K cells/well in antibiotic-free medium. The next day, cells were transfected with 200 ng total DNA/well (10:1:1 ratio as described for HASM1 cells) using FuGENE6 (1 ul/well) from Promega. Transfection complexes were replaced the following day with fresh complete medium and cells allowed to recover an additional 24 hr prior to 24 hr treatment and luciferase assay.

Luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega) and Infinite M1000 Plate Reader (Tecan) as previously described[49], with the following modifications for cells grown in 96-well plates: each well was lysed in 25 ul 1X lysis reagent and luminescence detected from 5 ul lysate combined with 40 ul of each assay reagent. Each experiment was performed in biologic sextuplet and repeated at least once with qualitatively similar results.

RNA Purification and Quantitative RT-PCR (qRT-PCR)

BEAS-2B cells were grown to confluence in 6-well tissue culture plates and treated with vehicle (DMSO) or TCDD (10 nM) for 4 hr. Wells were rinsed 3X in DPBS and then fresh medium containing 100 nM dex was added to all wells for an additional 4 hr. Cells were harvested in TRIzol and RNA purified using the PureLink RNA Mini Kit from Life Technologies prior to qRT-PCR, performed with normalization to *RPL19* as previously described[49]. Primers used for qRT-PCR are described in Supp Table 4.

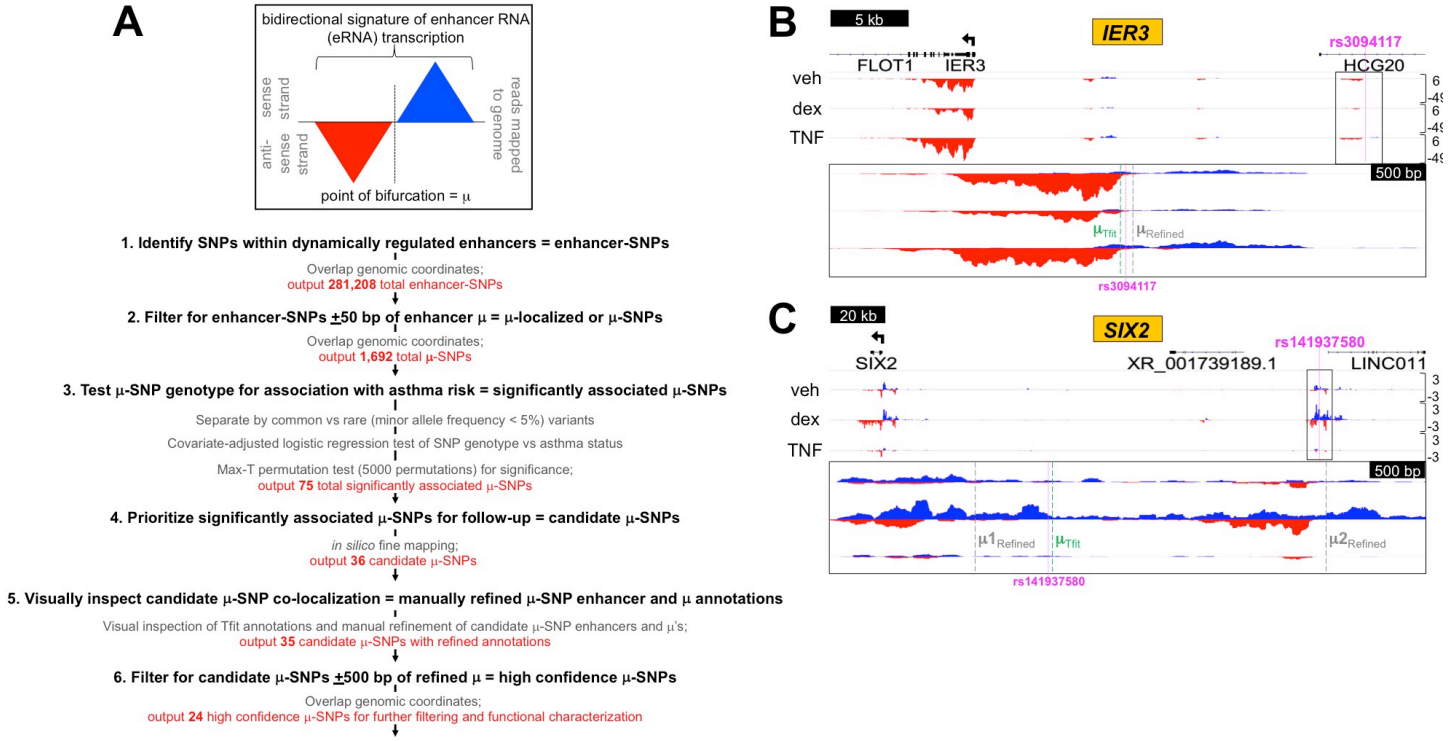

Figure 1. Pipeline for discovery of novel asthma-SNP associations based on proximity to μ and examples of μ-SNP colocalization. (A) Schematic of pipeline and filtering approaches to identify high confidence μ-SNPs associated with asthma. (**B-C**) GRO-seq tracks from BEAS-2B cells treated as indicated for 30 min and visualized in the Integrative Genomics Viewer (IGV) genome browser based on counts per million mapped reads (vertical scales). Blue indicates reads annotated to the sense strand while red indicates reads annotated to the antisense strand. The transcription start site (TSS) and direction of transcription are marked by arrows at the top of each screenshot. Magnified regions show the locations of μ originally calculated by Tfit (green) and following manual refinement (gray) relative to the indicated SNP within dynamically regulated enhancers containing high **(B)** or low **(C)** confidence μ-SNPs.

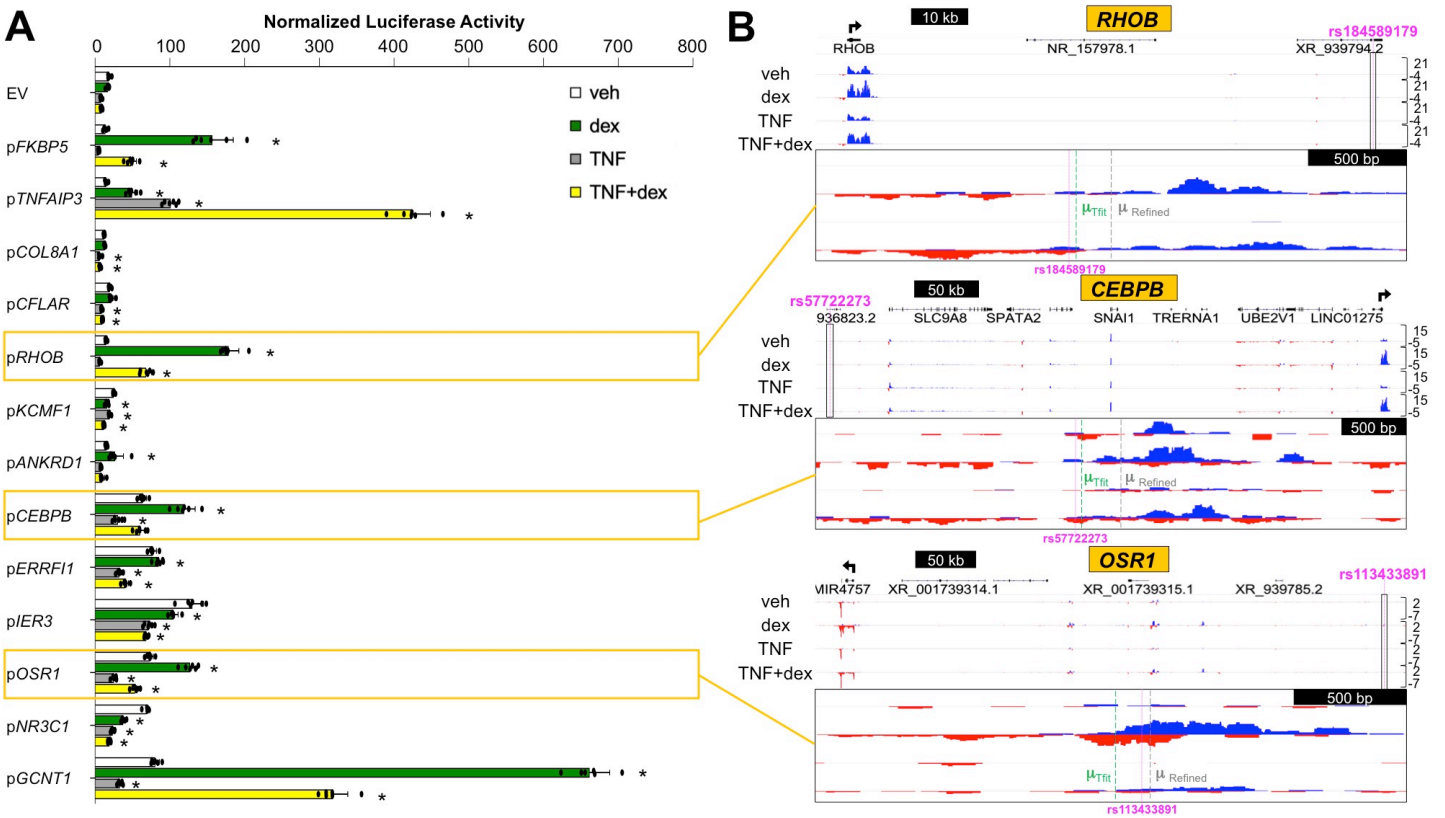

Figure 2. High confidence μ -SNP regions recapitulate dynamic eRNA regulation patterns in reporter assays. (A) Mean (±SD) normalized luciferase activity of indicated enhancer reporter constructs or empty vector (EV) control in BEAS-2B cells treated as indicated for 8 hr. p*FKBP5* and p*TNFAIP3* were included as canonical positive controls for dex and TNF responses. *p<0.05 vs same reporter+veh; one-way ANOVA corrected for multiple comparisons **(B)** GRO-seq tracks, as described for Figure 1, for select regions interrogated by enhancer reporter assay.

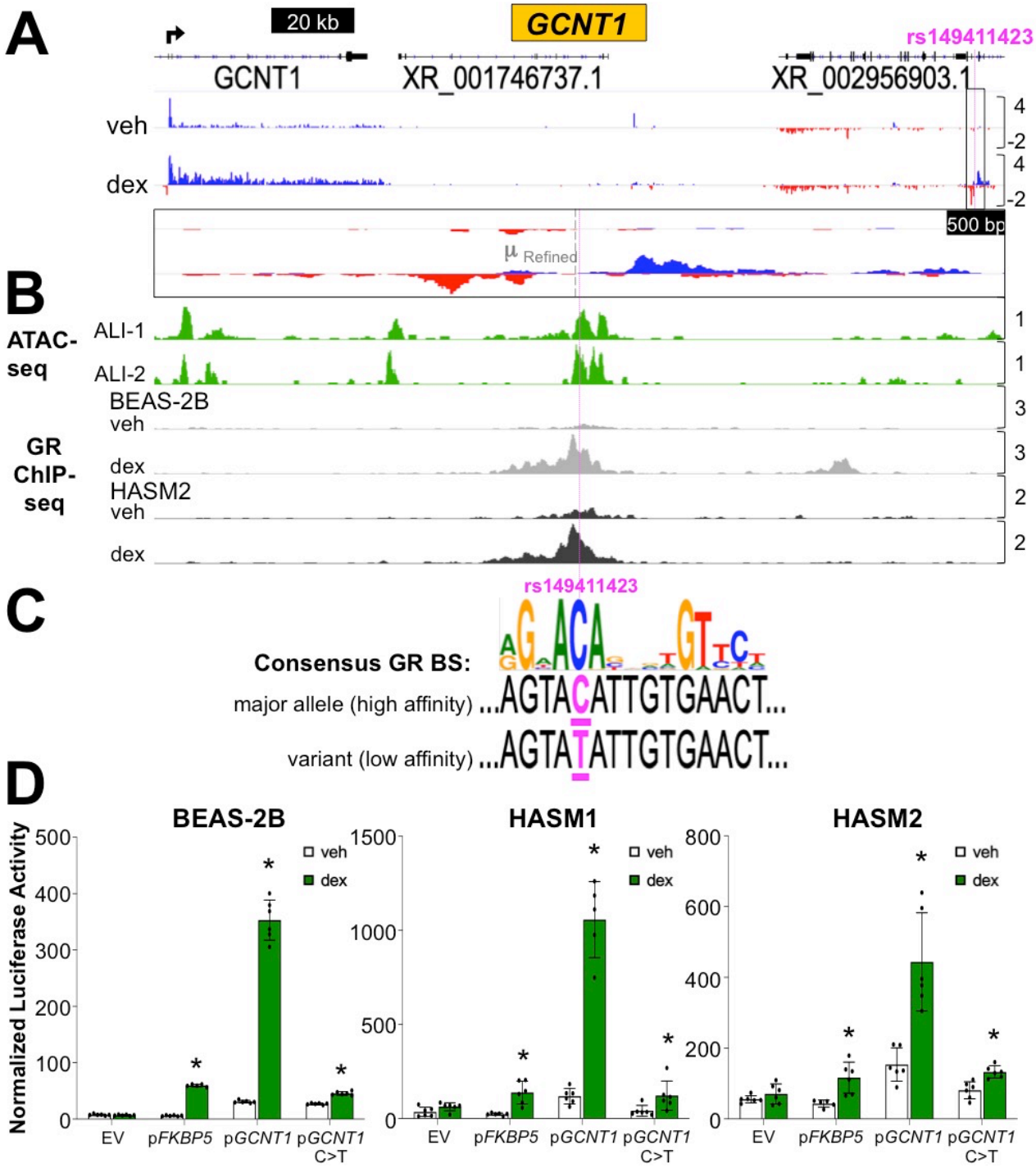

Figure 3. *rs14941123* disrupts a functional glucocorticoid response element that regulates induction of *GCNT1*. (A) GRO-seq tracks, as described for Figure 1, for *GCNT1* locus. **(B)** Aligned ATAC-seq tracks from primary human airway epithelial cells cultured at air-liquid interface (ALI; green) and glucocorticoid receptor (GR) ChIP-seq tracks from BEAS-2B (gray) and primary human airway smooth muscle (HASM) cells (black) ±1 hr dex visualized in the IGV browser based on counts per million mapped reads (vertical scales). **(C)** Consensus binding sequence logo (MatBase) for GR and sequence of match identified by MatInspector within the *GCNT1* enhancer harboring *rs14941123*. **(D)** Mean (±SD) normalized luciferase activity of WT or point mutation *GCNT1* reporter constructs in cells treated ±dex for 8 hr (BEAS-2B) or 24 hr (primary HASM). *p<0.05 vs same reporter+veh; unpaired t-test.

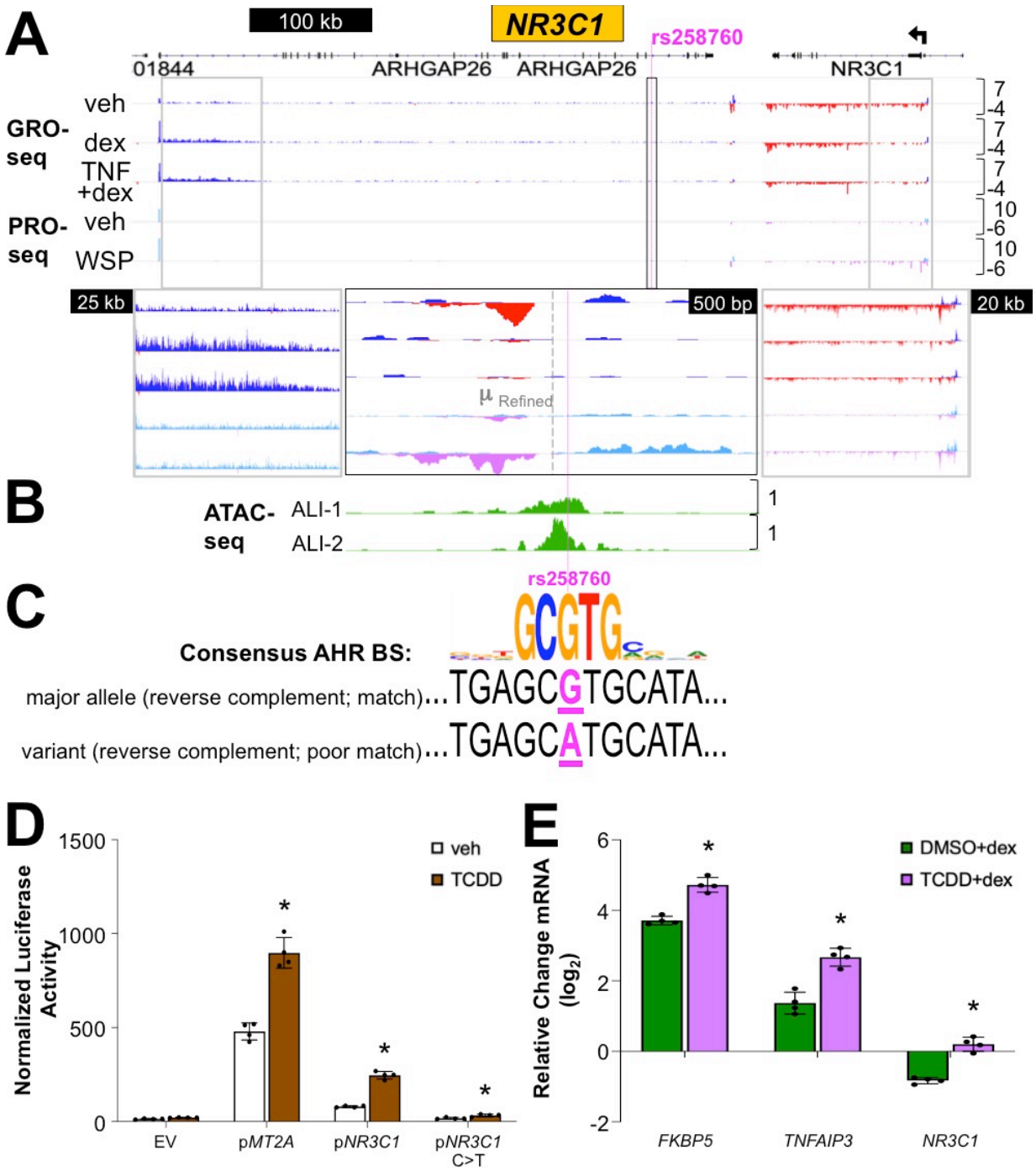

Figure 4. Expression and function of *NR3C1* are regulated by aryl hydrocarbon receptor (AHR) ligands through an AHR response element that is lost in the context of the *rs258760* variant allele. (A) IGV screenshots of BEAS-2B GRO-seq tracks aligned with PRO-seq tracks from BEAS-2B cells ±30 min wood smoke particles (WSP) at the *NR3C1* locus. PRO-seq reads annotated to the sense strand are light blue while those annotated to the antisense strand are light purple. **(B)** Aligned ATAC-seq tracks from primary human airway epithelial cells cultured at ALI. **(C)** Consensus binding sequence logo (MatBase) for the aryl hydrocarbon receptor (AHR) and sequence of match identified by MatInspector within the *NR3C1* enhancer containing *rs258760*. **(D)** Mean (±SD) normalized luciferase activity of WT or point mutation *NR3C1* reporter constructs in cells treated ±TCDD for 8 hr; p*MT2A* served as a positive control for TCDD responsiveness. *p<0.05 vs same reporter+veh; unpaired t-test. **(E)** qPCR analysis of dex-mediated gene regulation following 4 hr TCDD pre-treatment. Bars depict mean (±SD) C_T values on a $\log_2$ scale relative to DMSO+veh-treated cells. *p<0.05 vs same gene + DMSO+dex; unpaired t-test.

Table 1. Initial set of μ-SNPs (n=36) significantly associated with asthma risk in GERA cohort

SNP	Minor Allele Frequency (GERA)	Odds Ratio (Asthma risk)	SE	L95	U95	Differentially Regulated Enhancer Dataset Assoc. with μ-SNP	Permutation Analysis Pval
rs12506479	2.71E-01	1.05	0.02	1.01	1.09	TNF-induced	6.58E-03
rs1409773	6.30E-02	1.07	0.03	1.00	1.14	TNF-induced	3.64E-02
rs1454917	3.40E-01	1.04	0.02	1.00	1.07	dex-repressed	2.83E-02
rs258760	3.40E-01	0.95	0.02	0.92	0.99	dex-repressed	5.04E-03
rs3094117	2.72E-01	1.04	0.02	1.00	1.07	dex-repressed	3.69E-02
rs3129975	2.14E-01	1.08	0.02	1.04	1.12	dex-repressed	1.65E-04
rs3924229	8.19E-02	0.94	0.03	0.89	1.00	dex-induced	3.92E-02
rs58128597	7.54E-02	0.93	0.03	0.87	0.98	TNF-repressed	1.33E-02
rs6018492	8.64E-02	0.95	0.03	0.89	1.00	dex-induced	4.78E-02
rs60249091	1.82E-01	1.04	0.02	1.00	1.08	TNF-induced	4.33E-02
rs61289882	2.08E-01	1.05	0.02	1.01	1.09	TNF-induced	9.38E-03
rs113433891	1.97E-05	5.073	1.421	0.3132	82.16	dex-induced	2.34E-04
rs114363579	3.34E-04	2.464	0.3741	1.184	5.129	dex-repressed	9.30E-03
rs117861858	4.92E-05	10.42	0.9547	1.604	67.68	dex-repressed	1.13E-04
rs141243502	2.56E-04	0.1491	0.7984	0.03118	0.7132	dex-induced	6.50E-03
rs141870749	2.95E-05	9.824	1.473	0.5479	176.1	dex-induced	1.58E-04
rs141937580	1.97E-05	5.296	1.974	0.1107	253.4	dex-induced	2.40E-04
rs149411423	2.95E-05	13.2	1.692	0.4791	363.5	dex-induced	2.36E-04
rs150808839	1.17E-02	0.856	0.07654	0.7368	0.9945	dex-induced	4.29E-02
						TNF-induced	4.48E-02
rs150872583	2.95E-05	6.952	1.227	0.6272	77.06	dex-induced	1.59E-04
rs182978446	1.97E-05	10.79	1.515	0.5546	210	dex-induced	2.41E-04
rs184589179	2.72E-03	0.7231	0.1632	0.5251	0.9956	dex-induced	4.98E-02
rs186542091	8.92E-03	0.8214	0.08997	0.6887	0.9799	dex-repressed	3.29E-02
rs186639917	1.57E-04	5.43	0.6348	1.565	18.84	dex-induced	1.61E-03
rs186724791	1.57E-04	0.07542	1.231	0.006753	0.8423	dex-induced	2.24E-02
						TNF-induced	2.24E-02
rs187844646	1.97E-05	12.58	1.554	0.5983	264.5	dex-induced	2.36E-04
rs188187807	1.77E-04	2.729	0.5261	0.9733	7.652	dex-induced	4.20E-02
rs189641630	6.59E-04	2.155	0.2877	1.226	3.788	dex-induced	6.00E-03
rs57722273	2.38E-03	1.431	0.1563	1.053	1.943	dex-induced	2.13E-02
rs73731205	1.42E-03	0.5454	0.2431	0.3387	0.8783	dex-repressed	1.02E-02
rs76528672	7.87E-05	5.51	0.7271	1.325	22.92	dex-repressed	1.06E-02
rs76717004	3.09E-03	0.6658	0.161	0.4856	0.9128	dex-repressed	1.15E-02
rs78124271	1.48E-03	1.504	0.2011	1.014	2.23	dex-induced	4.03E-02
rs78292063	1.60E-03	1.551	0.1835	1.082	2.222	dex-induced	1.41E-02
rs78553489	2.07E-04	4.121	0.5667	1.357	12.51	TNF-repressed	4.71E-03
						dex-induced	6.21E-03
rs80028982	9.83E-06	3.088	1.423	0.1898	50.24	dex-induced	2.35E-04

References

1. Bønnelykke, K. & Ober, C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. *J Allergy Clin Immunol* **137**, 667-79 (2016).
2. Kim, K.W. & Ober, C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* **11**, 170-187 (2019).
3. Sleiman, P.M. *et al.* Variants of DENND1B associated with asthma in children. *N Engl J Med* **362**, 36-44 (2010).
4. Bønnelykke, K. *et al.* A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* **46**, 51-5 (2014).
5. Long, A., Bunning, B., Sampath, V., DeKruyff, R.H. & Nadeau, K.C. Epigenetics and the Environment in Airway Disease: Asthma and Allergic Rhinitis. *Adv Exp Med Biol* **1253**, 153-181 (2020).
6. Augustine, T., Al-Aghbar, M.A., Al-Kowari, M., Espino-Guarch, M. & van Panhuys, N. Asthma and the Missing Heritability Problem: Necessity for Multiomics Approaches in Determining Accurate Risk Profiles. *Front Immunol* **13**, 822324 (2022).
7. Vicente, C.T., Revez, J.A. & Ferreira, M.A.R. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology* **6**, e165 (2017).
8. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238-243 (2021).
9. Alsheikh, A.J. *et al.* The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genomics* **15**, 74 (2022).
10. Chen, Z., Boehnke, M., Wen, X. & Mukherjee, B. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda)* **11**(2021).
11. Ortega, V.E., Meyers, D.A. & Bleecker, E.R. Asthma pharmacogenetics and the development of genetic profiles for personalized medicine. *Pharmgenomics Pers Med* **8**, 9-22 (2015).
12. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
13. Demenais, F. *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet* **50**, 42-53 (2018).
14. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet* **11**, 424 (2020).
15. Grosveld, F., van Staalduinen, J. & Stadhouders, R. Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. *Annu Rev Genomics Hum Genet* **22**, 127-146 (2021).
16. Currin, K.W. *et al.* Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am J Hum Genet* **108**, 1169-1189 (2021).
17. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-7 (2010).
18. Dukler, N. *et al.* Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. *Genome Res* **27**, 1816-1829 (2017).
19. Mikhaylichenko, O. *et al.* The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32**, 42-57 (2018).
20. Sasse, S.K. *et al.* Nascent transcript analysis of glucocorticoid crosstalk with TNF defines primary and cooperative inflammatory repression. *Genome Res* **29**, 1753-1765 (2019).
21. Franco, H.L., Nagari, A. & Kraus, W.L. TNFalpha signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol Cell* **58**, 21-34 (2015).
22. Azofeifa, J.G. *et al.* Enhancer RNA profiling predicts transcription factor activity. *Genome Res* **28**, 334-44 (2018).
23. Kvale, M.N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1051-60 (2015).
24. Dahlin, A. *et al.* Large-scale, multiethnic genome-wide association study identifies novel loci contributing to asthma susceptibility in adults. *J Allergy Clin Immunol* **143**, 1633-1635 (2019).

25. Yoosefzadeh-Najafabadi, M., Eskandari, M., Belzile, F. & Torkamaneh, D. Genome-Wide Association Study Statistical Models: A Review. *Methods Mol Biol* **2481**, 43-62 (2022).
26. Gerber, A.N., Newton, R. & Sasse, S.K. Repression of transcription by the glucocorticoid receptor: A parsimonious model for the genomics era. *J Biol Chem* **296**, 100687 (2021).
27. Schmidt, S.F., Larsen, B.D., Loft, A. & Mandrup, S. Cofactor squelching: Artifact or fact? *Bioessays* **38**, 618-26 (2016).
28. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
29. Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933-42 (2005).
30. Kathryn, R. *et al.* Transcriptome-wide Association Study of Circulating IgE Levels Identifies Novel Targets for Asthma and Allergic Diseases. *medRxiv*, 2020.08.17.20176479 (2020).
31. Fonseca, K.L. *et al.* Deficiency in the glycosyltransferase Gcnt1 increases susceptibility to tuberculosis through a mechanism involving neutrophils. *Mucosal Immunol* **13**, 836-848 (2020).
32. Gally, F. *et al.* The MUC5B-associated variant rs35705950 resides within an enhancer subject to lineage- and disease-dependent epigenetic remodeling. *JCI Insight* **6**(2021).
33. Kadiyala, V. *et al.* Cistrome-based Cooperation between Airway Epithelial Glucocorticoid Receptor and NF-kappaB Orchestrates Anti-inflammatory Effects. *J Biol Chem* **291**, 12673-87 (2016).
34. Sasse, S.K. *et al.* Glucocorticoid Receptor ChIP-seq Identifies PLCD1 as a KLF15 Target that Represses Airway Smooth Muscle Hypertrophy. *Am J Respir Cell Mol Biol* (2017).
35. Sasse, S.K. *et al.* Response Element Composition Governs Correlations between Binding Site Affinity and Transcription in Glucocorticoid Receptor Feed-forward Loops. *J Biol Chem* **290**, 19756-69 (2015).
36. Young, T.M. *et al.* Identifying Toxicologically Significant Compounds in Urban Wildfire Ash Using In Vitro Bioassays and High-Resolution Mass Spectrometry. *Environ Sci Technol* **55**, 3657-3667 (2021).
37. Ghetu, C.C. *et al.* Wildfire Impact on Indoor and Outdoor PAH Air Quality. *Environ Sci Technol* **56**, 10042-10052 (2022).
38. Wentworth, G.R., Aklilu, Y.A., Landis, M.S. & Hsu, Y.M. Impacts of a large boreal wildfire on ground level atmospheric concentrations of PAHs, VOCs and ozone. *Atmos Environ (1994)* **178**, 19-30 (2018).
39. Sorg, O. AhR signalling and dioxin toxicity. *Toxicol Lett* **230**, 225-33 (2014).
40. Hubler, T.R. & Scammell, J.G. Intronic hormone response elements mediate regulation of FKBP5 by progestins and glucocorticoids. *Cell Stress Chaperones* **9**, 243-52 (2004).
41. Hauberg, M.E. *et al.* Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls. *Mol Psychiatry* **24**, 1685-1695 (2019).
42. Danko, C.G. *et al.* Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12**, 433-8 (2015).
43. Perkey, E. *et al.* GCNT1-Mediated O-Glycosylation of the Sialomucin CD43 Is a Sensitive Indicator of Notch Signaling in Activated T Cells. *J Immunol* **204**, 1674-1688 (2020).
44. Perera, F. & Nadeau, K. Climate Change, Fossil-Fuel Pollution, and Children's Health. *N Engl J Med* **386**, 2303-2314 (2022).
45. Xu, R. *et al.* Wildfires, Global Climate Change, and Human Health. *New England Journal of Medicine* **383**, 2173-2181 (2020).
46. Guarnieri, M. & Balmes, J.R. Outdoor air pollution and asthma. *Lancet* **383**, 1581-92 (2014).
47. Gupta, A. *et al.* Deconvolution of multiplexed transcriptional responses to wood smoke particles defines rapid aryl hydrocarbon receptor signaling dynamics. *J Biol Chem*, 101147 (2021).
48. Altonsy, M.O., Sasse, S.K., Phang, T.L. & Gerber, A.N. Context-dependent cooperation between nuclear factor kappaB (NF-kappaB) and the glucocorticoid receptor at a TNFAIP3 intronic enhancer: a mechanism to maintain negative feedback control of inflammation. *J Biol Chem* **289**, 8231-9 (2014).
49. Sasse, S.K. *et al.* The glucocorticoid receptor and KLF15 regulate gene expression dynamics and integrate signals through feed-forward circuitry. *Mol Cell Biol* **33**, 2104-15 (2013).