

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Sexually-dimorphic targeting of functionally-related genes in COPD

BMC Systems Biology 2014, **8**:118 doi:10.1186/s12918-014-0118-y

Kimberly Glass (rekr@channing.harvard.edu)
John Quackenbush (johnq@jimmy.harvard.edu)
Edwin K Silverman (reeks@channing.harvard.edu)
Bartolome Celli (bcelli@partners.org)
Stephen I Rennard (srennard@unmc.edu)
Guo-Cheng Yuan (gcyuan@jimmy.harvard.edu)
Dawn L DeMeo (redld@channing.harvard.edu)

Published online: 28 November 2014

ISSN 1752-0509

Article type Research article

Submission date 11 August 2014

Acceptance date 9 October 2014

Article URL <http://www.biomedcentral.com/1752-0509/8/118>

Like all articles in BMC journals, this peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to <http://www.biomedcentral.com/info/authors/>

Sexually-dimorphic targeting of functionally-related genes in COPD

Kimberly Glass^{1,2,3,*}
Email: rekr@channing.harvard.edu

John Quackenbush^{1,2,3}
Email: johnq@jimmy.harvard.edu

Edwin K Silverman^{3,4}
Email: reeks@channing.harvard.edu

Bartolome Celli⁴
Email: bcelli@partners.org

Stephen I Rennard⁵
Email: srennard@unmc.edu

Guo-Cheng Yuan^{1,2}
Email: gcyuan@jimmy.harvard.edu

Dawn L DeMeo^{3,4}
Email: redld@channing.harvard.edu

¹ Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA

² Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

³ Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁴ Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

⁵ Division of Pulmonary, Critical Care, Sleep and Allergy, University of Nebraska Medical Center, Omaha, NE, USA

* Corresponding author. Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Abstract

Background

There is growing evidence that many diseases develop, progress, and respond to therapy differently in men and women. This variability may manifest as a result of sex-specific

structures in gene regulatory networks that influence how those networks operate. However, there are few methods to identify and characterize differences in network structure, slowing progress in understanding mechanisms driving sexual dimorphism.

Results

Here we apply an integrative network inference method, PANDA (*Passing Attributes between Networks for Data Assimilation*), to model sex-specific networks in blood and sputum samples from subjects with Chronic Obstructive Pulmonary Disease (COPD). We used a jack-knifing approach to build an ensemble of likely networks for each sex. By adapting statistical methods to compare these network ensembles, we were able to identify strong differential-targeting patterns associated with functionally-related sets of genes, including those involved in mitochondrial function and energy metabolism. Network analysis also identified several potential sex- and disease-specific transcriptional regulators of these pathways.

Conclusions

Network analysis yielded insight into potential mechanisms driving sexual dimorphism in COPD that were not evident from gene expression analysis alone. We believe our ensemble approach to network analysis provides a principled way to capture sex-specific regulatory relationships and could be applied to identify differences in gene regulatory patterns in a wide variety of diseases and contexts.

Keywords

Network modeling, Gene regulation, Regulatory networks, Sexual-dimorphism, Chronic Obstructive Lung Disease

Background

Chronic respiratory diseases, including Chronic Obstructive Pulmonary Disease (COPD), are among the most likely causes of death in the United States; COPD ranks third only after heart disease and all forms of cancer combined [1]. In the past COPD was thought to primarily affect males, but in recent years the number of females with COPD has greatly increased, and currently more women die of COPD than men [2]. Some of the changing epidemiology is likely due to an increase in female cigarette use during the 1960s. However, current research also suggests biological causes for the apparent sexual-dimorphism in the disease, with women having a higher susceptibility [3-5], an overall more severe COPD course even with the same level of tobacco exposure [6], and an increase in severe symptoms at a younger age [2,7].

Investigating sex differences in disease is a critical area of investigation [8,9] and a wide number of diseases are known to affect men and women differently [10]. It has been noted that many sexually dimorphic features are likely not primarily due to genetic variation [11]. On the other hand, network-modeling of transcriptomes in model organisms has demonstrated sexually dimorphic higher-order gene interactions [12]. Consequently, systems-based approaches have great potential for exploring sex-differences in human traits [13,14].

In this study we leverage gene expression data from subjects with COPD to build sex-specific networks and investigate whether alterations in gene regulation might contribute to sexual-dimorphism in COPD. The methods described here are not limited to analysis of lung disease but are generalizable to other diseases that demonstrate sexually dimorphic characteristics.

Gene regulation involves the concerted activity of many distinct but non-independent regulatory mechanisms [12,14]. While no single experimental assay can fully capture the complexity of a given biological system, each provides information concerning a particular feature that influences, or results from, the state of a cell. Because of the complexity of gene regulatory processes, there is increased interest in modeling approaches capable of integrating multiple sources of regulatory information [15-19], and evidence suggests that these methods perform much better than those using individual data types in isolation [20].

Along these lines, we developed PANDA (*Passing Attributes between Networks for Data Assimilation*) [21], a “message passing” network inference method that integrates multiple types of genomic data. PANDA models information flow through networks under the assumption that both “transmitters” and “receivers” play active roles in modulating regulatory processes. In PANDA’s model of gene regulatory control, transcription factors are the transmitters and the receivers are their target genes. A set of initial connections linking transcription factors to potential downstream targets is inferred by mapping transcription factor binding sites (TFBS) to the genome. Gene expression profiles provide information on shared activation states for elements in the network and protein-protein interaction data provide information on co-regulatory processes. PANDA starts with initial networks and then uses the various data to iteratively update the network structures to more accurately fit the available information, until the process converges on a consensus regulatory network.

In applying PANDA, we construct phenotype-specific models and then look for variation in TF-target interactions (“edges”) to explore regulatory differences. One surprising result of applying PANDA in such a comparative analysis is that we are able to observe meaningful changes in regulatory patterns even for genes that are not differentially expressed [22].

The comparative analysis of phenotype-specific networks enabled by PANDA makes it particularly useful for studying sexual dimorphism in health and disease, where the absolute levels of gene expression in disease may be similar in male and female tissues but in which different regulatory processes may be active [14], including differences in transcription factor regulation in the presence of sex hormones [23,24]. If this is the case, identifying sexually dimorphic network variability and associating these network characteristics with specific disease processes can lead not only to a better understanding of the disease, but also to therapies optimized for men and women.

In this study we begin by analyzing blood and sputum gene expression data from subjects with COPD. We then explore whether gene regulatory networks, estimated using these data, contain sex-specific regulatory patterns. To do this we use PANDA to model “ensembles” of sex-specific regulatory networks in COPD and use these network ensembles to identify differences in network topologies that are associated with biological functions in a sex-specific manner. As opposed to analyzing or contrasting the properties of single networks, this ensemble approach to network analysis allows for the statistical quantification of network features. In this application, we demonstrate how Gene Set Enrichment Analysis (GSEA), which was originally designed to quantify the association of gene sets with differential expression changes, can be used to estimate the association of gene sets with alterations in

network features in light of this ensemble approach. However, more generally, our ensemble approach to network modeling allows for the principled investigation of differences in network properties using statistical tools developed for genomic and other high-dimensional data.

Results and discussion

Genes and gene sets are not strongly differentially-expressed between males and females with COPD in either blood or sputum

We obtained and analyzed gene expression data in sputum and blood samples from 132 subjects (44 females and 88 males) with COPD enrolled in the ECLIPSE study [25]. Affymetrix CEL files were downloaded and normalized using RMA [26], with probe-sets mapped to Entrez-gene IDs using a custom CDF [27]. An initial quality control of this data was performed by running a principal component analysis on the expression values for the 24 probe-sets located on the Y chromosome. A plot of the first versus the second principal component (Additional file 1: Figure S1A) indicates that although most samples cluster according to the sex ascribed in the phenotype data, there are six samples which do not cluster as expected. To minimize potential noise due to poor quality data or sex misclassification, we eliminated these six subjects from further consideration, leaving 42 female and 84 male COPD subjects with both sputum and blood gene expression data. A principal component analysis plot for these remaining samples, generated using expression information for genes located on the Y chromosome, is shown in Figure 1A; age, COPD Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage based on spirometry and pack-years of cigarette smoking for the corresponding subjects are shown in Figure 1B. We compared the age, COPD GOLD stage and pack-years of cigarette smoking between men and women and observe significant differences in age and pack-years but no significant difference in disease stage. This is consistent with previous observations that women often get similarly severe COPD at a younger age and with less smoke exposure [2,6,7] and highlights the importance of understanding the biologic features mediating sexual dimorphism in COPD. All subjects included in this analysis are former smokers.

Figure 1 Comparison of males and females with COPD using a standard differential-expression analysis approach. (A) A PCA analysis on expression data using genes located on the Y chromosome. Males and females cluster into two groups. **(B)** Covariate information for the 42 female and 84 male subjects included in the analysis. The statistical difference between sexes for age and pack-years was calculated using an unpaired two-sample t-test and the statistical difference between the sexes for GOLD stage was calculated by applying a chi-squared test to a two by three (sex by stage) contingency table. **(C)** The top most differentially-expressed genes based on a using an unpaired two-tailed t-test (after specifically excluding genes those on the sex chromosomes). Genes with higher average expression in female are colored pink and those with higher average expression in male are colored blue. **(D)** The results of a GSEA analysis looking for GO category differential-expression between males and females. The five most differentially-expressed GO categories in males and females in either sputum or blood are shown. Deeper shades of pink are used to denote greater significance in female while deeper shades of blue indicate greater significance in males. The scale is based on the $-\log$ FDR significance for categories enriched in females, resulting in positive values, and on the $+\log$ FDR significance for categories enriched in males, resulting in negative values. Note that the color-range extends

to an FDR significance of 10^{-3} in each sex even though the most significant categories found in this analysis only reach an FDR significance of around 10^{-2} . (E) GSEA “enrichment plots” for the two most significantly differentially-expressed GO categories according to the GSEA analysis in males and females in either sputum or blood.

For the remaining 126 subjects, a genome-wide differential expression analysis including the sex chromosome genes serves as a strong positive control on the expression data as the results identify many expected sex-related differences (Additional files 1: Figure S2 and Additional file 1: Tables S1–S2). We next excluded genes on the sex chromosomes and tested if autosomal genes were strongly differentially-expressed between males and females in either the sputum or blood samples, using an unpaired two-sample t-test. Using the sputum samples, no genes are significantly differentially expressed between males and females at an FDR less than 0.1. Only eight autosomal genes (listed in Figure 1C) are significantly differentially-expressed in blood between female and male COPD subjects at an FDR threshold of 0.1, suggesting that the removal of sex chromosome genes largely mitigates the sex-specific gene expression signal. Consequently, subsequent analyses exclude genes on the sex chromosomes.

Although very few autosomal *genes* are significantly differentially-expressed when comparing samples from males and females, it is still possible that *groups* of interacting genes, representing particular biological functions, might be *collectively* differentially-expressed in a sex-specific manner. We evaluated this possibility by performing Gene Set Enrichment Analysis (GSEA) [28]. We downloaded the java implementation of GSEA (<http://www.broadinstitute.org/gsea/>) and tested for the collective sex-specific differential expression for sets of genes annotated to Gene Ontology (GO) functional categories. GSEA uses a gene-by-sample table of expression values and information concerning sample features (in this analysis, subject sex) to rank genes based on their differential expression. It then uses this ranking to test if sets of genes (for example, those annotated to a particular GO term) have consistent changes in expression patterns, in our case, consistently higher expression levels in one sex compared to the other.

Figure 1D shows the five most differentially-expressed functional gene sets (hereafter, simply “functions” or “GO terms”) in males and females for both sputum (top panel) and blood (bottom panel). Several of the corresponding GSEA enrichment plots are presented in Figure 1E. Although the top functions are only marginally significant, both the blood and sputum analysis includes several interesting results. In sputum, the most differentially-expressed functions reach an FDR significance in the range of 0.01 to 0.15 and include GO terms such as “sterol biosynthetic process” and “steroid hydrolase activity”, which may play a role in sexual dimorphism. The GO functions more highly expressed in COPD blood samples in males compared to females include “cell killing” and “phagocytosis”, processes potentially related to COPD pathogenesis and severity [29,30].

Jack-knifing can be used to robustly estimate and compare regulatory networks

We also used a two-sample f-test to evaluate if the *variance* of any of the autosomal genes’ expression levels was significantly difference between females and males. We observe that in sputum samples over 1000 genes are differentially-variable at an FDR less than 0.1. We include these genes in Additional file 2. This observation, together with the plausible functional enrichment results, led us to next hypothesize that the differential *targeting* of

biological functions may play a critical role in sexual dimorphism in COPD. Specifically, it is possible that genes are differentially *co-expressed*, even if their overall average expression levels are not significantly different. If this differential co-expression is taken as evidence of differential co-regulation, as is done in PANDA, then potential transcription factors that are differential-targeting these genes can be identified (Additional file 1: Figure S3).

It has been suggested that regulatory relationships between transcription factors and genes likely have both stochastic and deterministic components, and thus may be better modeled by probability distributions as opposed to simple Boolean relationships [31,32]. Furthermore, in this application we recognized that differences in sample size between males and females could potentially influence predictions of regulatory network interactions. Motivated by this, we used PANDA [21] to calculate *ensembles* of networks based on jack-knifed sets of samples drawn from our initial male and female subject populations (Figure 2A).

Figure 2 Using ensembles of networks to robustly identify sex-specific interactions and their associated genes. (A) A cartoon summary of how we use PANDA to build ensembles of networks using a jack-knifing approach to resample the original expression data multiple types. (B-D) Volcano plots of the difference in mean edge weight across two ensembles of networks compared to the p-value of the difference in the edge weight distributions in the two ensembles. Comparisons include (B) female versus male sputum networks, (C) female versus male blood networks, (D) female versus male “random” networks. Edges identified as “different” in each comparison are shown as either pink (female-specific) or blue (male-specific). (E-G) Venn diagrams showing the overlap in genes targeted by the female-specific (pink) or male-specific (blue) edges. Note that a gene can be targeted by both a male-specific and a female-specific edge, but by different upstream transcription factors. There is a high level of overlap in the genes targeted by the identified sex-specific edges in both the sputum and blood networks. (H-J) A hypergeometric probability was used to determine the significance of overlap in male-specific genes with genes annotated to GO categories, and female-specific genes with genes annotated to GO categories. The top five categories enriched in the males and females for each comparison are shown.

Specially, As an input to PANDA, we constructed transcription-factor target networks using position-weight-matrices for 130 TFs recorded in the Jaspar database [33], mapping these to the promoter regions, defined as $[-750,+250]$ base-pairs around the transcription start site. We also include information regarding physical protein-protein interactions between human transcription factors [34]. To build ensembles of networks, we used a “jack-knife” [35], randomly selecting ten samples without replacement to create 400 gene expression data sets, 100 for each of four sample sets (blood-female, blood-male, sputum-female, sputum-male). We then used PANDA to infer networks for each expression data set. As a negative control, we also created a version of the sputum expression data with a permutation of gene labels, and built sex-specific ensembles of networks for this randomized data.

This jack-knifing approach ensures that the predicted network edges are not strongly influenced by any one subject, as each network in our ensembles represents an estimate of the cellular regulatory network for a subset of the relevant samples. It also helps us regularize differences in sample size between the sexes as each of the reconstructed networks contains information from the same number of subjects. Further, our male and female ensembles each include one hundred networks, giving us the power to quantify the statistical properties of the estimated regulatory edges, something that would have been difficult or impossible had we simply estimated a single network for each sex and tissue-type combination. Although the

jack-knifing approach does not allow us to directly model covariates (for example, differences in COPD severity or smoking histories), it helps mitigate their effect on the network predictions by modeling a distribution of networks, which are, on average, representative of the population, but whose variance likely represents the contribution of other factors.

We used an un-paired two-sample t-test to quantify differences in the distributions of predicted edge-weights between the sex-specific network ensembles. We also averaged the predicted edge weight across the networks in each ensemble, excluded edges with low average weights (<0) and, for the remaining edges, determined the difference in these average edge weight values between the ensembles. Figure 2B-D shows volcano plots of the difference in the average of each edge's weight between the ensembles being compared, versus the FDR significance in the difference of edge weight distributions based on the t-test. We immediately observe that edge differences in the "random" volcano plot are not nearly as strong as those in the sputum and blood volcano plots; however, there are some differences, including edges that are "significantly" different according to the t-test. Consequently in this following network edge analysis we use a more stringent FDR cutoff than we did with the gene expression analysis.

We used a combination of the difference (absolute value >0.25), significance based on the t-test ($\text{FDR} < 10^{-5}$) and average edge weight (>0) to select differentially-called edges for each ensemble comparison. Female- and male-specific edges are shown in pink and blue, respectively, in Figures 2B-D. These criteria were chosen such that each sex-specific subnetwork contains edges that are both likely to be real (based on a positive edge weight) as well as different, both at an absolute and at a statistical level. The cutoff values themselves were selected such that each subnetwork contains between one and five percent of all possible edges, which may be close to an expected network density. We applied these same cutoffs to the "random" volcano in order to quantify the level of false positives in the differential subnetwork edge calls. Although there are likely false-positive edges in our identified subnetworks, for the selected cut-offs there are approximately 2.4 and 9.4 times more differentially-called edges in the sputum and blood volcanos compared to the random volcano, respectively. We note that this randomization control also illustrates that statistical differences calculated by contrasting various network properties should be viewed primarily as a rank-ordering as opposed to a true significance level.

We determined the genes targeted by these sex-specific edges and present the results as Venn diagrams (Figure 2E-G). Many genes (5389 in sputum and 8133 in blood) are targeted in both male and female subnetworks, although the network models indicate the regulation is governed by different transcription factors. This may partially explain why we previously observed only minimal differential gene expression patterns between the sexes; our network results suggest that although genes may be similarly expressed in both sexes, this is mediated by a distinct set of transcriptional regulators.

To assess whether the genes targeted in only one sex-specific subnetwork and not the other might be associated with specific biological functions, we used Fisher's exact test to evaluate the enrichment of GO categories in these genes and observe some functional enrichment (Figure 2H-J). The signal appears to be strongest for the genes uniquely targeted in a sex-specific manner in the sputum-derived networks (Figure 2H); the sputum samples may be biologically "closer" to the disease as a lung source sample and may represent cellular process most likely to be associated with COPD.

Network ensembles uncover differential-targeting patterns in men and women with COPD

We recognize that there are significant limitations to studying functional enrichment in a context that relies upon somewhat arbitrary thresholds in order to define differential subnetworks (Figure 2B-J). Firstly, this type of approach can be sensitive to the cutoffs used, opening the opportunity for potentially biased results when not used with caution. Additionally, selecting genes based on whether they are or are not targeted in a pair of networks ignores any *relative* level of differential targeting. Specifically, we observe a high level of overlap in target genes when comparing male and female subnetworks (see Figure 2E-G); however, there are multiple instances when a gene is targeted by many transcription factors in one subnetwork but by a much smaller number, or even a single TF in the other. Although we excluded these commonly targeted genes in the analysis shown in Figure 2E-J, one could imagine they might play a significant role in sex-specific differences in COPD.

Motivated to overcome these limitations, we next used the ensembles of networks generated by PANDA in a manner analogous to how we used the expression data to evaluate differential-enrichment of GO functions between the sexes. We previously observed that some sets of functionally-related genes are weakly differentially-expressed (Figure 1D); here we wish to address a similar, but distinctly different question within the network context. Namely, are sets of functionally-related genes differentially-*targeted*? In other words, do a set of functionally-related genes tend to have an increase (or decrease) in regulatory targeting in one sex-specific regulatory network context compared to another?

In this analysis, instead of sets of *expression samples* associated with disease state and sex, we have sets of *regulatory networks*. Specifically, we have one hundred corresponding representative networks for each set of expression samples, and therefore one hundred predicted scores for each edge in those networks. Figure 3A shows a heat map of those scores for the male and female sputum networks. Some edges have consistently higher predicted edge weights in the male networks while others have consistently higher predicted edge weights in the female networks. We would like to relate these differences in network structure to differences in the regulation of biological functions.

Figure 3 Visualization of edge weight and the in- and out-degree of genes and TFs in ensembles of sputum networks. (A) Edge weights for every possible transcription factor to gene interaction, where each row represents an edge, and each column represents one of the networks produced in the jack-knifing approach. Rows are ordered based on the t-statistic comparing the edge weight values and each row is Z-score normalized for visualization purposes only. (B) The in-degree, defined as the sum of all incoming edge weights, for each gene in the PANDA network reconstruction. Genes (rows) are ordered based on the t-statistic comparing the gene in-degree distributions in the two ensembles of networks (columns). Again, rows are Z-score normalized only for visualization purposes. (C) The twenty-five most differentially-targeted genes, identified as having the most significant difference in in-degree in the male compared to the female ensemble of networks. Both the significance of the differential-targeting and the level of differential-expression is shown. (D) The out-degree, defined as the sum of all outgoing edges, for each transcription factor in the reconstructed networks. Rows represent transcription factors and are again ordered based on the t-statistic comparing the distribution of in-degree values of the transcription factor in the two ensembles of networks. (E) The ten most differentially-targeting transcription factors, and

their level of differential-expression. The majority of the differentially-targeted genes and differential-targeting transcription factors are not differentially-expressed.

To begin to address this question, within each of our sex-specific PANDA predicted networks, we assigned every gene a score based on its “in-degree”, which is defined as the sum of the weights of all edges pointing to that gene. Figure 3B shows the in-degree values side-by-side for the male and female sputum networks. We sorted genes in this figure based on the statistical difference in the in-degree values between the two network ensembles, as measured by an unpaired two-sample t-test. As with the edges, we observe that some genes are consistently much more highly targeted in the male networks, while others are consistently much more highly targeted in the female networks. The twenty-five most differentially-targeted genes, based on the t-test comparison, are shown in Figure 3C. As a control for this analysis we also reconstructed one hundred networks built after permuting the sex-labels of the subjects (Additional file 1: Figure S4A). We observe that the differential-targeting observed for these genes is much greater than expected by chance.

Our calculated in-degree values give an indication of how heavily a gene is targeted in a given network. Edge-weights predicted by PANDA correspond to how likely a given regulatory interaction is to exist and edges that represent either activating or repressing interactions can have similarly high weights. Consequently, genes with relatively higher degrees are not necessarily “more activated,” they may in fact be repressed (if they are highly targeted by more repressors than activators), or neither (if they are equally targeted by both activators and repressors). Therefore, a change in a gene’s degree between two sets of networks is not necessarily related to either an increase or decrease in its expression level, but instead suggests changes in its regulatory control. Consistent with this framework, even the most strongly differentially-targeted genes do not appear to be strongly differentially-expressed (Figure 3C). We therefore suggest that these differences in gene targeting likely represent a sexually-dimorphic disease-related re-wiring of the cellular network and that understanding the biological implications of these structural changes may provide insight into the mechanisms driving disease morphology and lead to suggestions for sex-specific therapies.

We also calculated the “out-degree” of TFs in these networks, or the sum of the weights of all edges pointing from a TF, and show the results in Figure 3D-E. As before, we observe strong sex-specific differences in targeting patterns, even though the TFs themselves are not differentially-expressed. These results suggest that differences in regulatory patterns in the absence of strong differential expression exist around the regulating TFs as well as the regulated genes. Thus the sex differences we observe appear to be strongest at the level of the network “edge” and not necessarily in the individual “node” (gene and TF) states.

Biological functions are strongly associated with sexually-dimorphic targeting in COPD subjects

Our analysis suggests that although there is little difference in gene expression levels between males and females with COPD in either blood or sputum, there are likely different regulatory mechanisms associated with and potentially mediating the disease state. If this is true, one would expect that alterations in network structure should be concentrated around genes representing particular functional classes representing changes in the *mechanisms of activation*, rather than downstream changes in gene expression. Therefore, next we sought to identify sexually dimorphic *differentially-targeted* functions. We created “gene-by-network”

tables for each ensemble of networks, where the values are the in-degrees of the genes (the level of targeting identified by PANDA) in each of our predicted networks. We then ran GSEA using these in-degree values instead of expression to evaluate if functionally-related sets of genes gain or lose targeting.

Running GSEA on differential gene-degree leads to some striking results (Figure 4A). First, despite the lack of strong differential-*expression* noted previously, directly comparing male versus female *networks* using this enrichment method reveals strong patterns of differential-*targeting*, with many functions that have significantly ($FDR < 0.01$) more targeting in the female compared to the male networks (Figure 4A). Differential-targeting of these functional categories is absent in networks reconstructed after permuting the sex-labels (Additional file 1: Figure S4B). Furthermore, the results are highly consistent when comparing female and male networks built using either the sputum or blood samples (although there is overall greater enrichment for differential-targeting of functions in the sputum). In contrast, repeating the analysis using networks constructed from “random” expression data shows no strong differential-targeting patterns.

Figure 4 Sexually-dimorphic targeting of biological functions in Sputum and Blood networks. (A) All GO categories significantly differentially-targeted ($FDR < 0.01$) using a GSEA-type approach to compare gene targeting in male and female networks derived from either sputum or blood expression data. Many functional categories have genes that appear to be much more highly targeted in the female networks compared to the male networks. There is a high level of agreement between the differential-targeted GO categories in both the sputum and blood networks, but the enrichment disappears in the “random” networks. (B) The ten most differentially-targeted pathways enriched in the female and the male sputum networks.

Closer inspection of the differentially-targeted functions shows many to be highly-related based on their biological role and gene content. Figure 4B shows the ten most differentially-targeted functions in females and males in sputum. A closer inspection of the expression levels of the genes annotated to these top functional categories shows that they appear to be associated with disease stage (Additional file 1: Figure S5), supporting their relevance to COPD. The pathways most significantly targeted in men are related to type I interferon, which has previously been implicated in the sexual dimorphism in response to viral infections (drivers of COPD exacerbations) [36,37] and in autoimmune diseases [38]. They are also consistent with previous observations that immune functions are enriched in male COPD-associated genes [39]. The pathways more highly targeted in women are all related in some way to mitochondrial function, which has previously been implicated in the modulation and development of lung disease [40,41]. Cigarette smoking has also been shown to change mitochondrial morphology [42] and abnormal mitochondrial function is described in patients with COPD [43,44].

Because of its maternal inheritance [45,46], the mitochondria has long been associated with sex-differences. Sex hormones play an important role in controlling mitochondrial biogenesis and activities [47-50]. In neuronal cells ER-beta is localized in the mitochondria and mediates mitochondrial vulnerability to oxidative damage [51,52]; it also impairs mitochondrial oxidative metabolism in mesothelioma [53]. Interestingly, estrogen receptors are reduced in the mitochondria of epithelial cells from asthmatic lungs [54]. In addition, multiple peroxisome proliferator-activated receptors (PPARs), a class of nuclear hormone receptor proteins, have lower expression levels in COPD patients. This activity corresponds to lower

expression levels of the PPAR- γ co-activator PGC-1 α [55], a key regulator of energy metabolism [56] and an inducer of mitochondria biogenesis [57]. Thus differential-targeting of mitochondrial functions is consistent both with known biology concerning sexual-dimorphism and COPD.

We have performed two analyses to confirm that the strong differential-targeting of biological functions we observe in these networks is not a consequence of our specific approach. First, we repeated the ensemble network reconstruction on the sputum expression data, but modified our sampling technique to match covariates between each selected set of ten female and ten male samples; the conclusions of this covariate-matched analysis are nearly identical to what we observe with the random sampling (Additional file 1: Figure S6). Secondly, we ran (1) one hundred GSEA differential-expression analyses, one for each set of ten versus ten expression samples, and (2) one hundred GSEA differential-targeting analyses, one for each female versus male network reconstructed from these samples. Across these analyses we again observe consistently strong differential-targeting of many biological functions (Additional file 1: Figure S7).

Transcription factors mediate differential-targeting patterns in COPD

To gain a better appreciation for the network-level patterns that might be driving the identified functional alterations, we constructed a gene-by-TF matrix of the t-statistic values associated with the differences in edge weights predicted for the female compared to the male sputum networks and performed a complete-linkage hierarchical clustering using a Pearson correlation coefficient distance (Figure 5A). The resulting heatmap, where the rows are genes and the columns are transcription factors, shows clear patterns involving sets of transcription factors differentially-targeting sets of genes in the female and male networks. Given these results, we next sought to identify if particular transcription factors might be mediating the differential-targeting of biological functions between men and women.

Figure 5 Transcription-factor differential-targeting of biological functions. (A) A hierarchical clustering of the t-statistic associated with differential edge weight between ensembles of female and male sputum networks. Each point in the matrix represents the t-statistic of an individual edge extending from a transcription factor (column) to a gene (row) (B-C) The statistical enrichment of GO categories in genes differentially-targeted by a transcription factor between male and female (B) sputum and (C) blood networks. All categories significantly (FDR < 0.01) differentially-targeted by at least one of the transcription factors, in the given tissue-type, is shown. The columns (transcription factors) are ordered identically to the hierarchical clustering in (A) and we observe a strong correlation with the transcription-factor differential-targeting of GO categories in (B). Although there is some similarity in the transcription-factor differential-targeting of these functional sets of genes in the sputum and blood networks, there is overall less enrichment in the blood comparison.

For each jack-knife iteration PANDA calculates an edge weight for every possible transcription factor to gene interaction representing the likelihood that the TF regulates that target gene. We used this information to design TF-specific gene-by-network tables. We ran GSEA on these TF-specific tables to evaluate if any functions are more strongly targeted by an individual TF in one of our ensembles of networks compared to the other. The results of the female versus male comparison in both sputum and blood are shown in Figure 5B-C, with the transcription factors shown in the same order as in the hierarchical clustering and each

row representing a biological function found to be enriched ($FDR < 0.01$) when contrasting at least one set of male or female TF-specific edges. We find more than 1000 GO functions differentially-targeted between the sexes by at least one transcription factor in sputum, and almost 900 in blood. As with the gene in-degree analysis we once again see much stronger differential-targeting of functions in the sputum network comparison relative to the blood network comparison.

Disease-specific regulators of sexually-dimorphic functional targeting

In order to better interpret this information, we focused on our previously-identified ten most differentially targeted functions (see Figure 4B) and present the TF-specific GSEA results in Figure 6A. We see overall consistency between the blood and sputum sexually-dimorphic targeting of these functions by individual transcription factors. However, a handful of transcription factors appear to have opposite patterns in the sputum and the blood networks.

Figure 6 Identifying disease-specific drivers of sexually-dimorphic functional targeting.

(A) Sputum (top panel) and blood (bottom panel) transcription-factor specific enrichment for differential-targeting of the top five GO functions identified in either males or females in Figure 4B. (B) A distribution of the similarity between differential-targeting patterns of transcription factors in sputum and blood network, as measured using the Spearman correlation. A red line indicates the cutoff used to identify transcription factors that have opposite sex-specific regulatory patterns in sputum compared to blood networks. (C-G) Plots comparing individual transcription factors' sex-specific differential-targeting of these GO functions (five in female, filled shapes, and five in male, hollow shapes) in the sputum versus the blood networks.

One limitation of directly comparing data from men and women with COPD is that without healthy controls it is unclear whether the systemic changes and high level of consistency we observe in the blood and sputum network analyses are important for sex-related differences in the disease or are a consequence of normal sex differences in cellular regulation. However, we reasoned that the sputum networks should be “closer” to lung disease, and thus transcription factors that are regulating biological functions in sputum but not in blood may be the most important drivers of sex-specific *and* disease-specific functional regulation. Therefore, to partially address our lack of healthy controls, we next directly compared the transcription-factor specific differential-targeting of functions in the sputum versus the blood networks.

We quantified differences in transcription-factor level targeting of the ten functions in Figure 6A by calculating, for each transcription factor, the Spearman correlation between the significance levels in the sputum sex-specific network comparison and the significance levels in the blood sex-specific network comparison. A distribution of these correlation values is shown in Figure 6B.

Most transcription factors have a high positive correlation value, indicating that they are increasing/decreasing their targeting of these biological functions between men and women similarly in both sputum and blood networks. Some of this sexually-dimorphic targeting may be related to COPD, however, it is also possible, since this behavior was observed in both sputum and blood samples, that it is a consequence of normal sex-differences. On the other hand, there is a relatively smaller subset of transcription factors – those with negative correlation coefficients – whose sexually-dimorphic targeting of these important functions is

opposite in the sputum and blood networks. We indicate the 23 transcription factors with correlation less than -0.4 by arrows in Figure 6A.

The transcription factors most differentially-targeting these key functions between sputum and blood, based on our correlation measure, include the HAND1::TCFE2A complex, FOXF2, PAX4, the MYC::MAX complex, and SOX5 (Figures 6C-G). Both FOXF2 and SOX5 have been implicated in COPD or lung biology and it is interesting that we observe them in this sex-specific context. For example, FOXF2 has been shown to quantitatively increase binding upon smoke exposure in female mice [58] and modulates the expression of lung genes [59]. SOX5 is a candidate for COPD susceptibility and important for lung development [60].

A network model for sex-specific targeting of functionally-related genes in COPD

The GSEA analysis we have performed based on the differential-targeting of genes is clearly very powerful and has led to the identification both of potential biological functions targeted in a sexually-dimorphic manner in COPD as well as several transcriptional regulators that may be mediating those differences. One strength of this analysis is that it relies upon characterizing network differences based on relative changes in targeting patterns. However, in doing so it also ignores the actual strength of predicted network interactions. In other words, if a gene is more targeted in one ensemble of networks relative to the other, that gene is highly implicated in the GSEA analysis, even if its input edges have low absolute edge-weight values predicted across all the networks in both ensembles. It is unlikely that the systemic differential-targeting of functions we see across our panel of transcription factors in Figure 6A actually corresponds to multiple strong regulatory interactions from every one of them.

To better appreciate the relationship between likely regulatory interactions and the results of our functional analysis, we next visualized subnetworks based on the female-specific and male-specific edges we previously identified (Figures 2B-C). In order to interpret our functional results in this regulatory network context we identified sex-specific edges that extend between the 23 disease-specific transcription factors and genes annotated to the top differentially-targeted functions. We illustrate the resulting subnetworks in Figure 7. Edges and genes are colored pink or blue based on whether they were identified as part of the female or male networks or functions, respectively.

Figure 7 Illustrations of core subnetworks of sex-specific regulation in COPD. (A) The sputum subnetwork, defined by sputum sex-specific edges (as shown in Figure 2B) that extend between the identified disease-specific drivers (Figure 6B) and genes annotated to one of the top five differentially-targeted GO categories in ensembles of networks built based on sputum expression data (Figures 4B and 6). **(B)** The blood subnetwork, defined by blood sex-specific edges (Figure 2C) that extend between the identified disease-specific drivers (Figure 6B) and genes annotated to top five GO categories identified as differentially-targeted between male and female networks (Figures 4B and 6). Edges and genes are colored pink or blue based on whether they were identified as part of the female or male networks or functions, respectively.

We observe that some transcription factors, such as CREB1 and ZFX target distinct sets of genes in both the male and female sputum networks. ZFX is the X-linked version of a protein

that plays a role in molecular sex determination [61], so it may not be surprising that we found sex-specific differences in its regulatory patterns. However, it has also been implicated in lung cancer [62,63]. Similarly CREB is over-expressed in many cancers, including lung cancer [64], and, interestingly, has been shown to interact with the estrogen receptor and to have age and sex- dependent expression patterns in the human brain [65].

Several transcription factors dominate in one sex compared to the other. For example, the MYC::MAX complex appears to primarily target genes annotated to functions enriched in the female-specific sputum network (but not in the blood network) while SOX5 targets genes in the male-specific sputum network. USF1, in particular, appears to be a “hub” transcription factor for the female functionally related-genes in the sputum networks.

USF1 both regulates and interacts directly with estrogen receptor (ER) in a protein complex [66], which may explain its female-specific activity. Estrogen has also been shown to induce USF1 to bind to the regulatory regions of several genes [67-69]. USF1 is involved in the cross-talk between hypoxia-related elements such as Aryl hydrocarbon receptor (AHR) and the estrogen receptor, inhibiting the former [70-72]. This relationship may be important in sex-specific COPD biology as AHR has previously been identified as potentially important for sex-specific differences in lung cancer [73].

Sex-specific effects of USF1 have been noted previously [74,75]. Consistent with our findings, it has been reported that in mouse liver, male gene signatures are enriched for functions such as immune response while female signatures are enriched in functions such as oxidoreductase activity and mitochondrion [75]. ChIP of USF1 in HepG2 cells also indicates that it regulates nuclear mitochondrial genes [76]. Most interestingly, however, is that fact that USF1 has been shown to bind to the promoter and mediate the expression of PGC-1 α [77,78] which, as we previously noted, is an important regulator of mitochondrial biogenesis [57], and, along with several PPARs, has been shown to be expressed at lower levels in the skeletal muscle of COPD patients [55]. Therefore, it is not unreasonable to suppose that USF1, as indicated by our PANDA network analysis, may be an important mediator of mitochondrial-activity in a sexually-dimorphic manner in patients with COPD.

Conclusions

In this study we identified functionally related sets of genes that are strongly differentially-targeted between men and women with COPD. Our results suggest that sexual dimorphism in features of COPD may be a consequence of the re-wiring of cellular networks around particular biological pathways, especially those involved in mitochondrial function and energy metabolism, leading to differences in COPD in men and women. Although these functions have previously been implicated in COPD, little is known about their disease- and sex-specific regulation. In addition, despite the fact that there is a large body of research concerning the structural features of *individual* regulatory networks [79-82], quantifying *differences* in network features is relatively understudied and there are few systematic approaches for characterizing variability in gene targeting. In our analysis we contrasted networks and identified functionally related sets of genes that are strongly *differentially-targeted* between men and women with COPD.

One of our most striking findings was clear differential-targeting patterns in the absence of similarly compelling differential-expression. Several potential biological mechanisms may

play a role in mediating this differential targeting. One possibility is that multiple transcription factors compete for the same binding site upstream of a given target gene, but which one primarily regulates that gene is dependent on the cellular context (for example a change in protein abundance or conformation in response to sex hormones). Another possibility is that several transcription factors have potential binding sites upstream of a gene, but in females certain sites are inactive (for example through an epigenetic factor or a mutation) and in males others are inactive.

Using a network-based approach we were able to identify potential sex- and disease-specific transcriptional regulators of these biological functions, the most striking of which was USF1. Although USF1 has previously been implicated both in the regulation of nuclear mitochondrial genes and in sexual-dimorphism, its specific role in COPD is largely unknown and our findings are an important step in beginning to understand its potential importance. Curiously, an increase or decrease in *overall* out-degree by transcription factors between male and female networks did not always directly correspond to differential-targeting of particular biological *functions* between male and female networks. For example USF1 had an overall higher out-degree in male networks (Figure 3E), yet it also had increased targeting of mitochondrial functions in female networks (Figures 6 and 7). This highlights the importance of interpreting network measures within a functional context.

As with any computational analysis, there are limitations in our investigation that result from the underlying data we used; for example the number of genes included on the expression array may affect the comprehensiveness of the information incorporated in the model. One limitation in our specific application is that, although we found many sex-specific regulatory features, the sputum and blood expression data we used was only collected from individuals with COPD, and thus we lacked truly “normal” controls—this is a crucial direction for future research. However, by focusing on sex differences we observed just in the sputum networks and not the blood networks, we believe our findings are likely to represent sex-specific network alterations that are important for COPD. We also used a covariate-free model to evaluate differential-expression in order to be consistent with our subsequent regulatory network analysis, which does not directly model the role of covariates. It is therefore possible that in addition to the sex-specific regulatory changes we observe, there may also be gene expression differences between men and women with COPD that are simply not captured using a covariate-free approach. However, we suggest that is equally likely that similar outcomes in gene expression are mediated by distinct sets of transcriptional regulators. For example, it is reasonable to imagine that sex hormones (such as estrogen), which we only modeled in our network through receptor binding sites, might change the functions of some transcription factors (for example USF1) in other ways, requiring cells to respond and differentially rewire the effected portion of their regulatory network in order to maintain viability. In this case the overall expression profile of the cells might be similar, but the factors mediating that response could be vastly different.

Genomic assays, such as gene expression data, provide a snapshot of the state of a cell and most widely used analysis approaches identify differences in *individual genes* by collectively comparing groups of samples. We believe one limitation of gene-centered approaches, especially in the context explored here, is due to the fact that individual genes do not define the biological processes that drive cell states, but that phenotypic alterations are better characterized by networks of interactions linking genes. In contrast, our *network approach*, although complementary to differential gene expression analysis, highlights fundamentally different aspects of sex-specific biology. Namely, that a gene, or a collection of genes

involved in a biological function, may be similarly *expressed* in both men and women, but this expression may be regulated by different upstream factors. Understanding how the *targeting* of biological functions is distinct between sexes in COPD helped to elucidate potentially sexually-dimorphic mechanisms of the disease, an endeavor with relevance for both sex-specific diagnostics and therapeutics. Differential targeting of biological pathways is likely not limited to sex-specific disease features, and we believe the methods we employ here will be widely applicable to better understanding other biological systems and diseases.

Methods

Building ensembles of PANDA networks using a jack-knife

We used PANDA [21] to integrate expression information with transcription factor motif and protein-interaction data. In our analysis we parsed the expression data by sex, employed a jack-knifing procedure and ran PANDA multiple times to construct sets of sex-specific, genome-wide transcriptional regulatory networks. The specifics of how we processed the input data and reconstructed the PANDA network models are included below:

Expression data

We obtained the CEL data files for 264 expression experiments performed on blood and sputum samples collected from 132 individuals and profiled using Affymetrix HGU 133 plus2 microarrays. We RMA-normalized the expression data in R [26,83], and mapped probes to Entrez-gene IDs using a custom CDF [27]. These data include 18960 probes sets, mapping to 18895 unique genes (based on Hugo Gene Symbols). 15820 of these genes are also included in our motif scan (see below), including 651 on the sex chromosomes. After an initial PCA analysis investigating the clustering of samples based on the expression of genes on the Y chromosome, we excluded genes on the sex chromosomes and removed expression samples for 6 individuals who did not cluster correctly according to sex. We used expression data for the remaining 126 individuals (42 females and 84 males) and 15169 genes when constructing sex- and tissue-specific genome-wide regulatory networks. We also created a “random” version of the sputum expression data by permuting autosomal gene labels.

Motif data

We obtained position weight matrixes (PWM) for 130 core vertebrate transcription factor motifs from JASPAR [84,85]. To identify the target locations for each motif, each candidate sequence S was given a motif score equal to $\log [P(S|M)/P(S|B)]$, where $P(S|M)$ is the probability of observing sequence S given motif M , and $P(S|B)$ is the probability of observing sequence S given the genome background B . We modeled the distribution of these motif scores by randomly sampling the genome 10^6 times. Motif sites with a significance level of $p < 10^{-5}$ and that fell within the promoter region ($[-750, 250]$ base-pairs around the transcriptional start site) of one of the genes measured on our expression arrays were used to defined as an edge between a motif and that gene in our regulatory network prior.

It is important to note that although when building our primary network models we did not include genes on the sex chromosome as potential targets, we did *not* remove motif information for sites bound by transcription factors encoded on the sex chromosomes (such as AR on chrX and SRY on chrY). We reasoned that since the motif sequences for these

transcription factors still exist in the regulatory regions of autosomal genes they can still be indicative of information about a target gene's local regulatory network structure.

PPI data

Interactions between human transcription factors were obtained from the supplemental material of [34]. We excluded interactions in this set when either transcription factor in the interaction did not directly match one of the motifs included in our regulatory network prior.

Reconstructing PANDA networks

To construct multiple sex- and tissue- specific network models, we selected ten subjects of the same sex at random, identified the sputum, blood and “random” expression data associated with these subjects, and used PANDA to, separately, integrate each of these three sample-sets of expression data with motif and protein-protein interaction data. We did multiple random selections of subjects of the same sex, constructing one hundred female-specific and one hundred male-specific networks for each tissue.

Identifying differentially-called edges between network ensembles

PANDA reports the probability that an edge exists between a transcription factor (i) and gene (j) in an estimated network (n) as a Z-score ($Z_{ij}^{(n)}$). To select edges that are differentially-called between male and female network ensembles, for each edge we calculated (1) its average edge-score across all networks in each of the two ensembles, (2) the difference between these average scores, and (3) used a t-test to evaluate the significance of the difference in edge-score distribution between the male and female network ensembles. We corrected this significance for multiple hypothesis testing. To select edges distinct between female and male network ensembles, we then selected edges that had an average edge score greater than zero in at least one of the ensembles, an absolute edge score difference of at least 0.25, and an FDR significance less than 10^{-5} .

Clustering network differences

In order to better appreciate large scale patterns in male/female sputum regulatory network differences, we performed a hierarchical clustering on a transcription factor by gene matrix populated with the t-statistic value of the corresponding network edge, calculated when comparing the distribution of predicted edge scores across the male versus female sputum network ensembles. Hierarchical clustering was done separately in each dimension using one minus the Pearson correlation as the distance metric and the “complete” linkage method.

GSEA

To run GSEA in a consistent manner on both gene expression and network regulatory data, we downloaded the java command line version of the program from www.broadinstitute.org/gsea/. We ran GSEA permuting gene set labels. Further, in order to ensure consistency between the GSEA analysis and the functional enrichment analysis we performed using Fisher's exact test (shown in Figure 2H-J), we ran the GSEA program using custom Gene Matrix Transposed (GMT) files that we constructed from human GO annotations downloaded from geneontology.org (access date: 02/02/13).

In analyzing the GSEA results, we consider the FDR p-values reported by GSEA. Specifically, we report enrichment in female categories based on the $-\log_{10}(\text{FDR})$ significance (resulting in positive values), and enrichment in male categories based on the $+\log_{10}(\text{FDR})$ significance (resulting in negative values).

Competing interests

Dr. Silverman reports grants from NIH, grants and other support from COPD Foundation, grants and personal fees from GlaxoSmithKline, personal fees from Merck and travel support from Novartis. Dr. Celli reports research activities for GlaxoSmithKline, Almirall, Novartis, Rorrest and Aeris and consultant activities for GlaxoSmithKline, Boehringer Ingelheim, Dey, Altana, Almirall, Pfizer and Rox. Dr. Rennard reports relationships (board membership/consultancy/lectures/research funding) since 2011 with AARC, American Board of Internal Medicine, Able Associates, Align2 Acton, Almirall, APT, AstraZeneca, American Thoracic Society, Beilenson, Boehringer Ingelheim, Chiesi, CIPLA, Clarus Acuity, CME Incite, COPDFoundation, Cory Paeth, CSA, CSL Behring, CTS Carmel, Dailchi Sankyo, Decision Resources, Dunn Group, Easton Associates, Elevation Pharma, FirstWord, Forest, GLG Research, Gilead, Globe Life Sciences, GlaxoSmithKline, Guidepoint, Health Advance, HealthStar, HSC Medical Education, Johnson and Johnson, Leerink Swan, LEK, McKinsey, Medical Knowledge, Medimmune, Merck, Navigant, Novartis, Nycomed, Osterman, Pearl, PeerVoice, Penn Technology, Pennside, Pfizer, Prescott, Pro Ed Communications, PriMed, Pulmatrix, Quadrant, Regeneron, Saatchi and Saatchi, Sankyo, Schering, Schlesinger Associates, Shaw Science, Strategic North, Summer Street Research, Synapse, Takeda, Telecon SC, and ThinkEquity. Drs. Glass, DeMeo, Yuan and Quackenbush declare no conflicts of interest.

Authors' contributions

KG and DLD conceptualized and designed the study; EKS, BC and SIR collected the data, KG, DLD, JQ and GY participated in analysis and statistical support. All authors contributed to the writing and editing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This project has been supported by R01 HL089438, R01 HL111759-01A1, R01 HL111759-02S1 and P01 HL105339. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ECLIPSE clinical Investigators — Bulgaria: Y. Ivanov, Pleven; K. Kostov, Sofia. Canada: J. Bourbeau, Montreal; M. Fitzgerald, Vancouver, BC; P. Hernandez, Halifax, NS; K. Killian, Hamilton, ON; R. Levy, Vancouver, BC; F. Maltais, Montreal; D. O'Donnell, Kingston, ON. Czech Republic: J. Krepelka, Prague. Denmark: J. Vestbo, Hvidovre. The Netherlands: E. Wouters, Horn-Maastricht. New Zealand: D. Quinn, Wellington. Norway: P. Bakke, Bergen. Slovenia: M. Kosnik, Golnik. Spain: A. Agusti, J. Sauleda, P. de Mallorca. Ukraine: Y. Feschenko, V. Gavrisyuk, L. Yashina, Kiev; N. Monogarova, Donetsk. United Kingdom: P. Calverley, Liverpool; D. Lomas, Cambridge; W. MacNee, Edinburgh; D. Singh, Manchester; J. Wedzicha, London. United States: A. Anzueto, San Antonio, TX; S. Braman, Providence, RI; R. Casaburi, Torrance CA; B. Celli, Boston; G. Giessel, Richmond, VA; M. Gotfried,

Phoenix, AZ; G. Greenwald, Rancho Mirage, CA; N. Hanania, Houston; D. Mahler, Lebanon, NH; B. Make, Denver; S. Rennard, Omaha, NE; C. Rochester, New Haven, CT; P. Scanlon, Rochester, MN; D. Schuller, Omaha, NE; F. Sciurba, Pittsburgh; A. Sharafkhaneh, Houston; T. Siler, St. Charles, MO; E. Silverman, Boston; A. Wanner, Miami; R. Wise, Baltimore; R. ZuWallack, Hartford, CT.

Steering Committee: H. Coxson (Canada), C. Crim (GlaxoSmithKline, USA), L. Edwards (GlaxoSmithKline, USA), D. Lomas (UK), W. MacNee (UK), E. Silverman (USA), R. Tal Singer (Co-chair, GlaxoSmithKline, USA), J. Vestbo (Co-chair, Denmark), J. Yates (GlaxoSmithKline, USA).

Scientific Committee: A. Agusti (Spain), P. Calverley (UK), B. Celli (USA), C. Crim (GlaxoSmithKline, USA), B. Miller (GlaxoSmithKline, USA), W. MacNee (Chair, UK), S. Rennard (USA), R. Tal-Singer (GlaxoSmithKline, USA), E. Wouters (The Netherlands), J. Yates (GlaxoSmithKline, USA).

References

1. Heron M: **Deaths: leading causes for 2010.** *Natl Vital Stat Rep* 2013, **62**(6):1–97.
2. Silverman EK, Weiss ST, Drazen JM, Chapman HA, Carey V, Campbell EJ, Denish P, Silverman RA, Celedon JC, Reilly JJ, Ginns LC, Speizer FE: **Gender-related differences in severe, early-onset chronic obstructive pulmonary disease.** *Am J Respir Crit Care Med* 2000, **162**(6):2152–2158.
3. Sin DD, Cohen SB, Day A, Coxson H, Pare PD: **Understanding the biological differences in susceptibility to chronic obstructive pulmonary disease between men and women.** *Proc Am Thorac Soc* 2007, **4**(8):671–674.
4. Han MK, Postma D, Mannino DM, Giardino ND, Buist S, Curtis JL, Martinez FJ: **Gender and chronic obstructive pulmonary disease: why it matters.** *Am J Respir Crit Care Med* 2007, **176**(12):1179–1184.
5. Sorheim IC, Johannessen A, Gulsvik A, Bakke PS, Silverman EK, DeMeo DL: **Gender differences in COPD: are women more susceptible to smoking effects than men?** *Thorax* 2010, **65**(6):480–485.
6. Martinez FJ, Curtis JL, Sciurba F, Mumford J, Giardino ND, Weinmann G, Kazerooni E, Murray S, Criner GJ, Sin DD, Hogg J, Ries AL, Han M, Fishman AP, Make B, Hoffman EA, Mohsenifar Z, Wise R: **Sex differences in severe pulmonary emphysema.** *Am J Respir Crit Care Med* 2007, **176**(3):243–252.
7. Hemsing N, Greaves L: **Women, environments and chronic disease: shifting the gaze from individual level to structural factors.** *Environ Health Insights* 2009, **2**:127–135.
8. Schiebinger L: **Scientific research must take gender into account.** *Nature* 2014, **507**(7490):9.
9. Pollitzer E: **Biology: cell sex matters.** *Nature* 2013, **500**(7460):23–24.

10. Kim AM, Tingen CM, Woodruff TK: **Sex bias in trials and treatment must end.** *Nature* 2010, **465**(7299):688–689.
11. Vige A, Gallou-Kabani C, Junien C: **Sexual dimorphism in non-Mendelian inheritance.** *Pediatr Res* 2008, **63**(4):340–347.
12. van Nas A, Guhathakurta D, Wang SS, Yehya N, Horvath S, Zhang B, Ingram-Drake L, Chaudhuri G, Schadt EE, Drake TA, Arnold AP, Lusk AJ: **Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks.** *Endocrinology* 2009, **150**(3):1235–1249.
13. Civelek M, Lusk AJ: **Systems genetics approaches to understand complex traits.** *Nat Rev Genet* 2014, **15**(1):34–48.
14. Arnold AP, Lusk AJ: **Understanding the sexome: measuring and reporting sex differences in gene systems.** *Endocrinology* 2012, **153**(6):2551–2555.
15. D’Haeseleer P, Liang S, Somogyi R: **Genetic network inference: from co-expression clustering to reverse engineering.** *Bioinformatics* 2000, **16**(8):707–726.
16. Guthke R, Moller U, Hoffmann M, Thies F, Topfer S: **Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection.** *Bioinformatics* 2005, **21**(8):1626–1634.
17. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA: **Combining location and expression data for principled discovery of genetic regulatory network models.** In *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing.* ; 2002:437–449.
18. Kato T, Tsuda K, Asai K: **Selective integration of multiple biological data for supervised network inference.** *Bioinformatics* 2005, **21**(10):2488–2495.
19. Youn A, Reiss DJ, Stuetzle W: **Learning transcriptional networks from the integration of ChIP-chip and expression data in a non-parametric model.** *Bioinformatics* 2010, **26**(15):1879–1886.
20. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R: **Gene regulatory network inference: data integration in dynamic models-a review.** *Biosystems* 2009, **96**(1):86–103.
21. Glass K, Huttenhower C, Quackenbush J, Yuan GC: **Passing messages between biological networks to refine predicted interactions.** *PLoS One* 2013, **8**(5):e64832.
22. Glass K, Quackenbush J, Spentzos D, Haibe-Kains B, Yuan GC: *A Network Model for Angiogenesis in Ovarian Cancer (submitted).* ; 2014.
23. Bjornstrom L, Sjoberg M: **Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes.** *Mol Endocrinol* 2005, **19**(4):833–842.
24. Heemers HV, Tindall DJ: **Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex.** *Endocr Rev* 2007, **28**(7):778–808.

25. Singh D, Fox SM, Tal-Singer R, Plumb J, Bates S, Broad P, Riley JH, Celli B: **Induced sputum genes associated with spirometric and radiological disease severity in COPD ex-smokers.** *Thorax* 2011, **66**(6):489–495.
26. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**(2):249–264.
27. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F: **Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.** *Nucleic Acids Res* 2005, **33**(20):e175.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545–15550.
29. Berenson CS, Kruzel RL, Eberhardt E, Sethi S: **Phagocytic dysfunction of human alveolar macrophages and severity of chronic obstructive pulmonary disease.** *J Infect Dis* 2013, **208**(12):2036–2045.
30. Taylor AE, Finney-Hayward TK, Quint JK, Thomas CM, Tudhope SJ, Wedzicha JA, Barnes PJ, Donnelly LE: **Defective macrophage phagocytosis of bacteria in COPD.** *Eur Respir J* 2010, **35**(5):1039–1047.
31. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell.** *Science* 2002, **297**(5584):1183–1186.
32. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, van Oudenaarden A: **Systematic identification of signal-activated stochastic gene regulation.** *Science* 2013, **339**(6119):584–587.
33. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B: **JASPAR: an open-access database for eukaryotic transcription factor binding profiles.** *Nucleic Acids Res* 2004, **32**(Database issue):D91–D94.
34. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, *et al*: **An atlas of combinatorial transcriptional regulation in mouse and man.** *Cell* 2010, **140**(5):744–752.
35. Wu CFJ: **Bootstrap and other resampling methods in regression analysis.** *Ann Stat* 1986, **14**(4):1261–1295.
36. Karnam G, Rygiel TP, Raaben M, Grinwis GC, Coenjaerts FE, Rensing ME, Rottier PJ, de Haan CA, Meyaard L: **CD200 receptor controls sex-specific TLR7 responses to viral infection.** *PLoS Pathog* 2012, **8**(5):e1002710.

37. Geurs TL, Hill EB, Lippold DM, French AR: **Sex differences in murine susceptibility to systemic viral infections.** *J Autoimmun* 2012, **38**(2–3):J245–J253.
38. Hughes GC: **Progesterone and autoimmune disease.** *Autoimmun Rev* 2012, **11**(6–7):A502–A514.
39. Faner R, Gonzalez N, Cruz T, Kalko SG, Agusti A: **Systemic inflammatory response to smoking in chronic obstructive pulmonary disease: evidence of a gender effect.** *PLoS One* 2014, **9**(5):e97491.
40. Aravamudan B, Thompson MA, Pabelick CM, Prakash YS: **Mitochondria in lung diseases.** *Expert Rev Respir Med* 2013, **7**(6):631–646.
41. Hara H, Araya J, Ito S, Kobayashi K, Takasaka N, Yoshii Y, Wakui H, Kojima J, Shimizu K, Numata T, Kawaishi M, Kamiya N, Odaka M, Morikawa T, Kaneko Y, Nakayama K, Kuwano K: **Mitochondrial fragmentation in cigarette smoke-induced bronchial epithelial cell senescence.** *Am J Physiol Lung Cell Mol Physiol* 2013, **305**(10):L737–L746.
42. Hoffmann RF, Zarrintan S, Brandenburg SM, Kol A, de Bruin HG, Jafari S, Dijk F, Kalicharan D, Kelders M, Gosker HR, Ten Hacken NH, van der Want JJ, van Oosterhout AJ, Heijink IH: **Prolonged cigarette smoke exposure alters mitochondrial structure and function in airway epithelial cells.** *Respir Res* 2013, **14**:97.
43. Puente-Maestu L, Perez-Parra J, Godoy R, Moreno N, Tejedor A, Gonzalez-Aragoneses F, Bravo JL, Alvarez FV, Camano S, Agusti A: **Abnormal mitochondrial function in locomotor and respiratory muscles of COPD patients.** *Eur Respir J* 2009, **33**(5):1045–1052.
44. Meyer A, Zoll J, Charles AL, Charloux A, de Blay F, Diemunsch P, Sibia J, Piquard F, Geny B: **Skeletal muscle mitochondrial dysfunction during chronic obstructive pulmonary disease: central actor and therapeutic target.** *Exp Physiol* 2013, **98**(6):1063–1078.
45. Wolff JN, Gemmell NJ: **Mitochondria, maternal inheritance, and asymmetric fitness: why males die younger.** *Bioessays* 2013, **35**(2):93–99.
46. Camus MF, Clancy DJ, Dowling DK: **Mitochondria, maternal inheritance, and male aging.** *Curr Biol* 2012, **22**(18):1717–1721.
47. Gavrilova-Jordan LP, Price TM: **Actions of steroids in mitochondria.** *Semin Reprod Med* 2007, **25**(3):154–164.
48. Yager JD, Chen JQ: **Mitochondrial estrogen receptors—new insights into specific functions.** *Trends Endocrinol Metab* 2007, **18**(3):89–91.
49. Klinge CM: **Estrogenic control of mitochondrial function and biogenesis.** *J Cell Biochem* 2008, **105**(6):1342–1351.

50. Capllonch-Amer G, Llado I, Proenza AM, Garcia-Palmer FJ, Gianotti M: **Opposite effects of 17-beta estradiol and testosterone on mitochondrial biogenesis and adiponectin synthesis in white adipocytes.** *J Mol Endocrinol* 2014, **52**(2):203–214.
51. Yang SH, Liu R, Perez EJ, Wen Y, Stevens SM Jr, Valencia T, Brun-Zinkernagel AM, Prokai L, Will Y, Dykens J, Koulen P, Simpkins JW: **Mitochondrial localization of estrogen receptor beta.** *Proc Natl Acad Sci U S A* 2004, **101**(12):4130–4135.
52. Yang SH, Sarkar SN, Liu R, Perez EJ, Wang X, Wen Y, Yan LJ, Simpkins JW: **Estrogen receptor beta as a mitochondrial vulnerability factor.** *J Biol Chem* 2009, **284**(14):9540–9548.
53. Manente AG, Valenti D, Pinton G, Jithesh PV, Daga A, Rossi L, Gray SG, O’Byrne KJ, Fennell DA, Vacca RA, Nilsson S, Mutti L, Moro L: **Estrogen receptor beta activation impairs mitochondrial oxidative metabolism and affects malignant mesothelioma cell growth in vitro and in vivo.** *Oncogenesis* 2013, **2**:e72.
54. Simoes DC, Psarra AM, Mauad T, Pantou I, Roussos C, Sekeris CE, Gratziou C: **Glucocorticoid and estrogen receptors are reduced in mitochondria of lung epithelial cells in asthma.** *PLoS One* 2012, **7**(6):e39183.
55. Remels AH, Schrauwen P, Broekhuizen R, Willems J, Kersten S, Gosker HR, Schols AM: **Peroxisome proliferator-activated receptor expression is reduced in skeletal muscle in COPD.** *Eur Respir J* 2007, **30**(2):245–252.
56. Liang H, Ward WF: **PGC-1alpha: a key regulator of energy metabolism.** *Adv Physiol Educ* 2006, **30**(4):145–151.
57. Austin S, St-Pierre J: **PGC1alpha and mitochondrial metabolism—emerging concepts and relevance in ageing and neurodegenerative disorders.** *J Cell Sci* 2012, **125**(Pt 21):4963–4971.
58. Tharappel JC, Cholewa J, Espandiari P, Spear BT, Gairola CG, Glauert HP: **Effects of cigarette smoke on the activation of oxidative stress-related transcription factors in female A/J mouse lung.** *J Toxicol Environ Health A* 2010, **73**(19):1288–1297.
59. Yang Z, Hikosaka K, Sharkar MT, Tamakoshi T, Chandra A, Wang B, Itakura T, Xue X, Uezato T, Kimura W, Miura N: **The mouse forkhead gene Foxp2 modulates expression of the lung genes.** *Life Sci* 2010, **87**(1–2):17–25.
60. Hersh CP, Silverman EK, Gascon J, Bhattacharya S, Klanderman BJ, Litonjua AA, Lefebvre V, Sparrow D, Reilly JJ, Anderson WH, Lomas DA, Mariani TJ: **SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development.** *Am J Respir Crit Care Med* 2011, **183**(11):1482–1489.
61. Han SH, Yang BC, Ko MS, Oh HS, Lee SS: **Length difference between equine ZFX and ZFY genes and its application for molecular sex determination.** *J Assist Reprod Genet* 2010, **27**(12):725–728.

62. Jiang M, Xu S, Yue W, Zhao X, Zhang L, Zhang C, Wang Y: **The role of ZFX in non-small cell lung cancer development.** *Oncol Res* 2012, **20**(4):171–178.
63. Zha W, Cao L, Shen Y, Huang M: **Roles of mir-144-ZFX pathway in growth regulation of non-small-cell lung cancer.** *PLoS One* 2013, **8**(9):e74175.
64. Sakamoto KM, Frank DA: **CREB in the pathophysiology of cancer: implications for targeting transcription factors for cancer therapy.** *Clin Cancer Res* 2009, **15**(8):2583–2587.
65. Paramanik V, Thakur MK: **Role of CREB signaling in aging brain.** *Arch Ital Biol* 2013, **151**(1):33–42.
66. deGraffenried LA, Hopp TA, Valente AJ, Clark RA, Fuqua SA: **Regulation of the estrogen receptor alpha minimal promoter by Sp1, USF-1 and ERalpha.** *Breast Cancer Res Treat* 2004, **85**(2):111–120.
67. Xing W, Archer TK: **Upstream stimulatory factors mediate estrogen receptor activation of the cathepsin D promoter.** *Mol Endocrinol* 1998, **12**(9):1310–1321.
68. Ikeda K, Inoue S, Orimo A, Tsutsumi K, Muramatsu M: **Promoter analysis of mouse estrogen-responsive finger protein (efp) gene: mouse efp promoter contains an E-box that is also conserved in human.** *Gene* 1998, **216**(1):155–162.
69. Dillner NB, Sanders MM: **Upstream stimulatory factor (USF) is recruited into a steroid hormone-triggered regulatory circuit by the estrogen-inducible transcription factor delta EF1.** *J Biol Chem* 2002, **277**(37):33890–33894.
70. Dasgupta C, Chen M, Zhang H, Yang S, Zhang L: **Chronic hypoxia during gestation causes epigenetic repression of the estrogen receptor-alpha gene in ovine uterine arteries via heightened promoter methylation.** *Hypertension* 2012, **60**(3):697–704.
71. Jiang B, Mendelson CR: **USF1 and USF2 mediate inhibition of human trophoblast differentiation and CYP19 gene expression by Mash-2 and hypoxia.** *Mol Cell Biol* 2003, **23**(17):6117–6128.
72. Wang F, Samudio I, Safe S: **Transcriptional activation of cathepsin D gene expression by 17beta-estradiol: mechanism of aryl hydrocarbon receptor-mediated inhibition.** *Mol Cell Endocrinol* 2001, **172**(1–2):91–103.
73. Mollerup S, Berge G, Baera R, Skaug V, Hewer A, Phillips DH, Stangeland L, Haugen A: **Sex differences in risk of lung cancer: expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts.** *Int J Cancer* 2006, **119**(4):741–744.
74. Fan YM, Hernesniemi J, Oksala N, Levula M, Raitoharju E, Collings A, Hutri-Kahonen N, Juonala M, Marniemi J, Lyytikainen LP, Seppala I, Mennander A, Tarkka M, Kangas AJ, Soininen P, Salenius JP, Klopp N, Illig T, Laitinen T, Ala-Korpela M, Laaksonen R, Viikari J, Kähönen M, Raitakari OT, Lehtimäki T: **Upstream Transcription Factor 1 (USF1)**

allelic variants regulate lipoprotein metabolism in women and USF1 expression in atherosclerotic plaque. *Sci Rep* 2014, **4**:4650.

75. Wu S, Mar-Heyming R, Dugum EZ, Kolaitis NA, Qi H, Pajukanta P, Castellani LW, Lusis AJ, Drake TA: **Upstream transcription factor 1 influences plasma lipid and metabolic traits in mice.** *Hum Mol Genet* 2010, **19**(4):597–608.

76. Rada-Iglesias A, Ameer A, Kapranov P, Enroth S, Komorowski J, Gingeras TR, Wadelius C: **Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders.** *Genome Res* 2008, **18**(3):380–392.

77. Irrcher I, Ljubicic V, Hood DA: **Interactions between ROS and AMP kinase activity in the regulation of PGC-1alpha transcription in skeletal muscle cells.** *Am J Physiol Cell Physiol* 2009, **296**(1):C116–C123.

78. Yoboue ED, Devin A: **Reactive oxygen species-mediated control of mitochondrial biogenesis.** *Int J Cell Biol* 2012, **2012**:403870.

79. Restrepo J, Ott E, Hunt B: **Characterizing the dynamical importance of network nodes and links.** *Phys Rev Lett* 2006, **97**(9):094102.

80. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298**(5594):824–827.

81. Barabasi AL, Albert R: **Emergence of scaling in random networks.** *Science* 1999, **286**(5439):509–512.

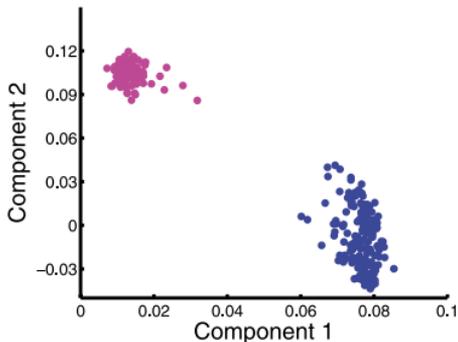
82. Girvan M, Newman ME: **Community structure in social and biological networks.** *Proc Natl Acad Sci U S A* 2002, **99**(12):7821–7826.

83. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I: **The Edinburgh human metabolic network reconstruction and its functional analysis.** *Mol Syst Biol* 2007, **3**:135.

84. Stormo GD: **DNA binding sites: representation and discovery.** *Bioinformatics* 2000, **16**(1):16–23.

85. Wasserman WW, Sandelin A: **Applied bioinformatics for the identification of regulatory elements.** *Nat Rev Genet* 2004, **5**(4):276–287.

A PCA on chrY Gene Expression



B Covariate Information

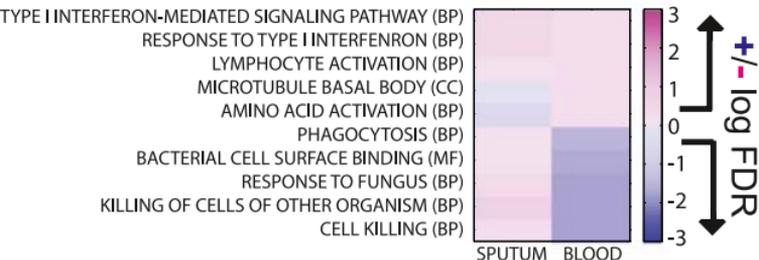
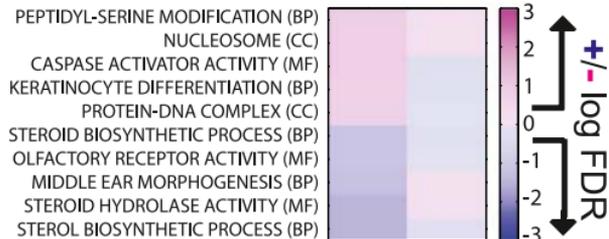
	Female	Male	P-value
Number of Subjects	42	84	X
Average Age	62.5	66.3	1.4×10^{-4}
Average Pack Years	37.6	53.6	4.5×10^{-3}
Gold Stage	II	23	0.14
	III	18	
	IV	1	

C Blood Differential Expression

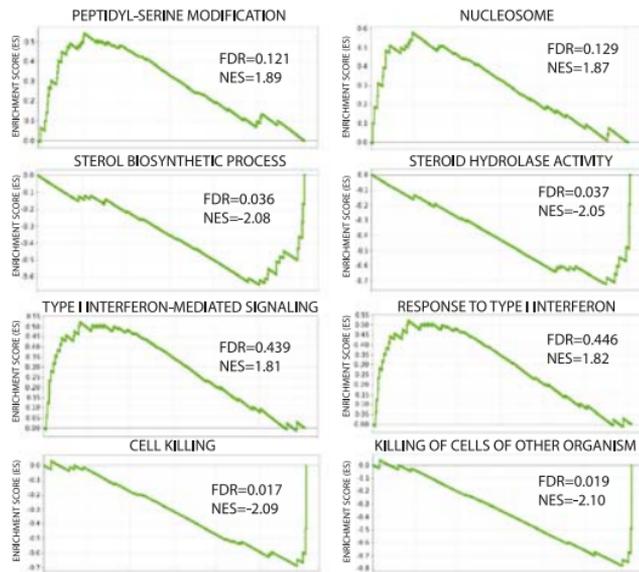
Gene	FDR	Fold-Change
STK32B	6.79E-04	0.934
FRG2C	9.68E-03	0.949
OAF	0.010	0.964
ASGR1	0.048	0.967
KDELC1	0.048	1.030
TSHZ3	0.048	0.958
CD96	0.061	1.034
S1PR3	0.082	0.950

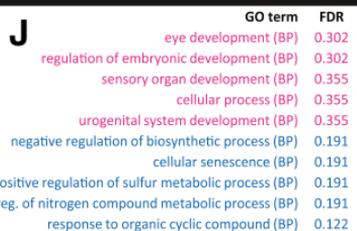
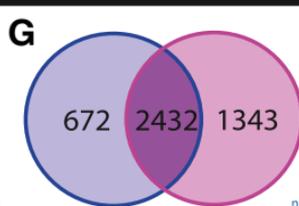
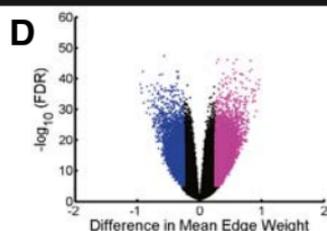
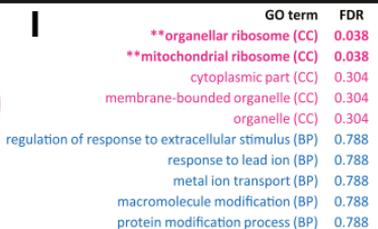
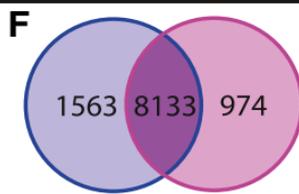
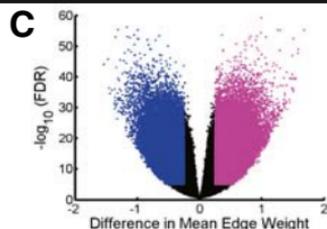
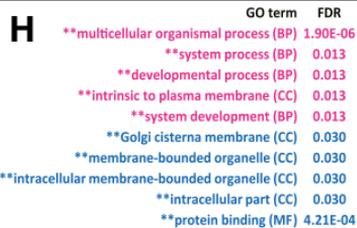
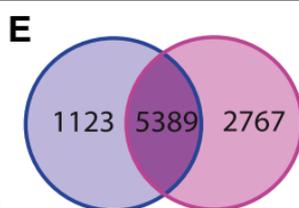
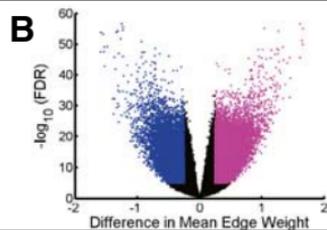
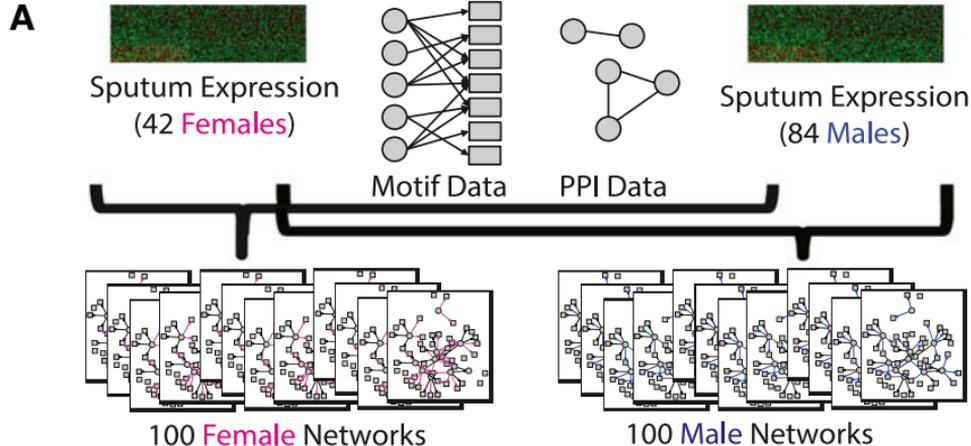
* No significant genes found when comparing sexes using sputum expression data.

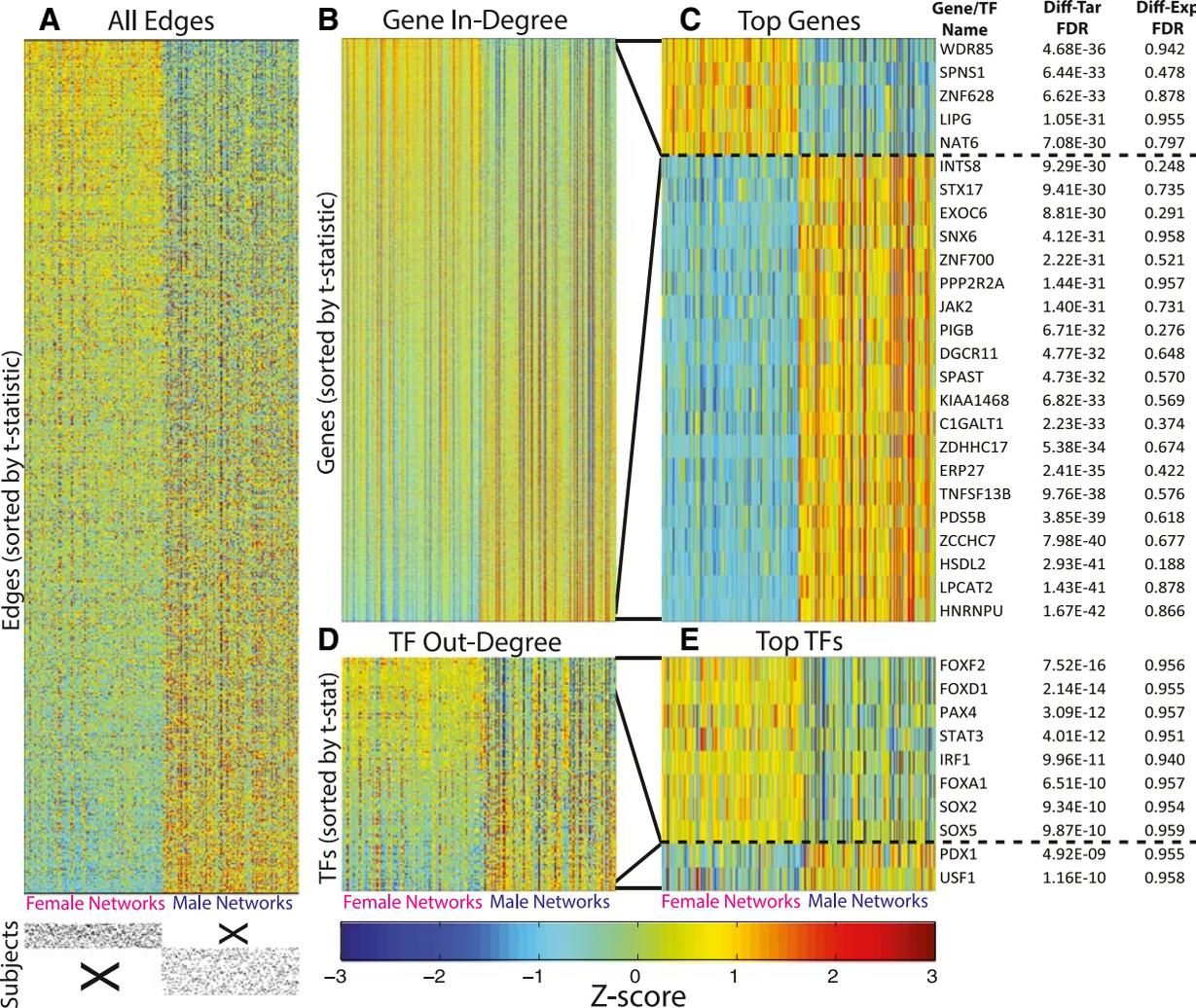
D

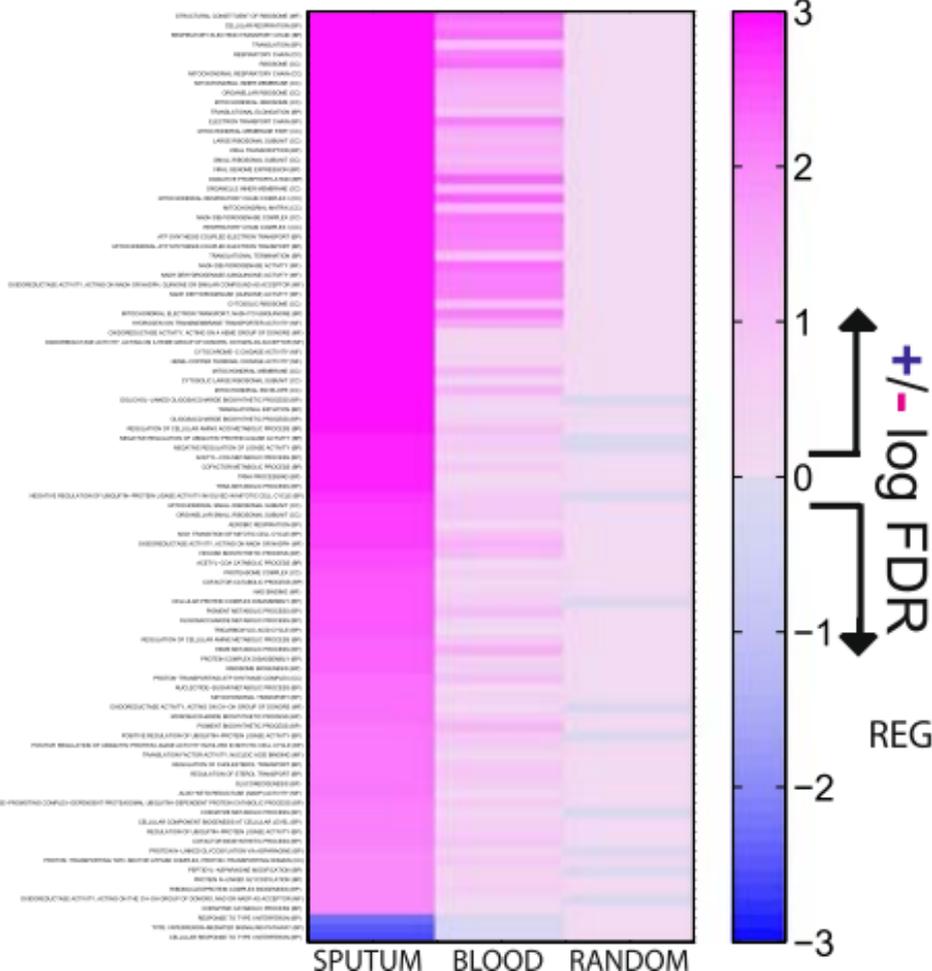
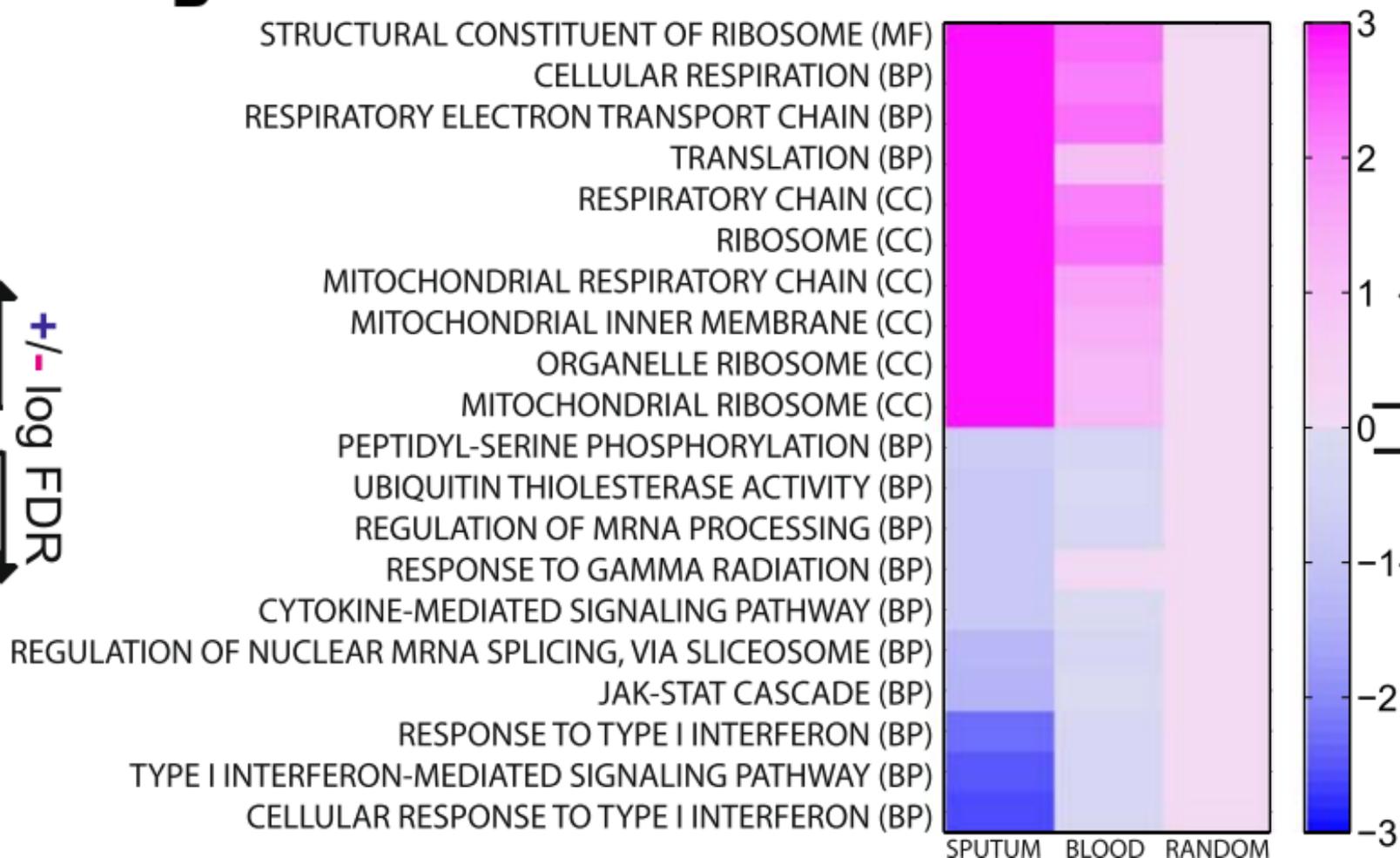


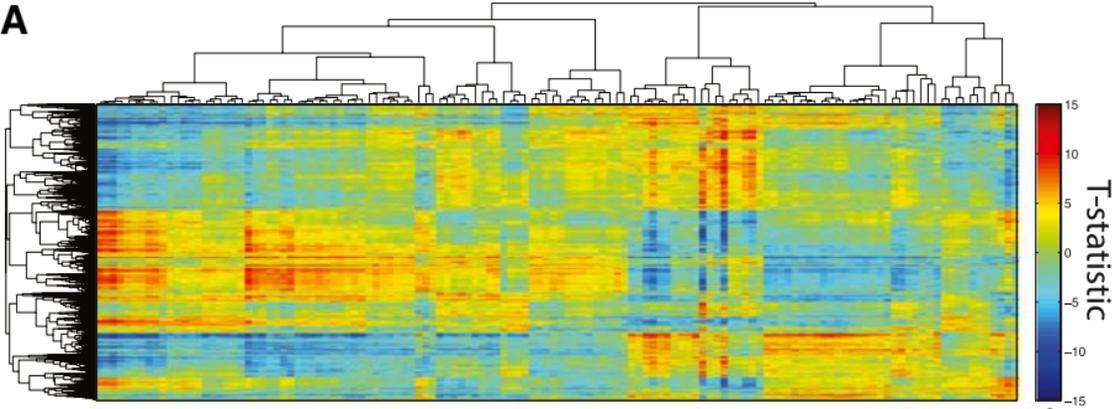
E



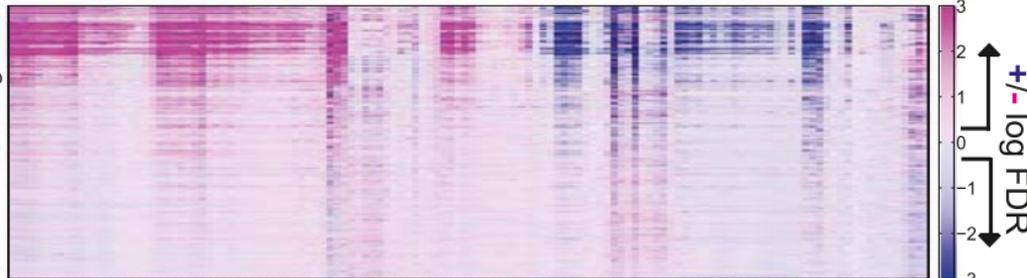




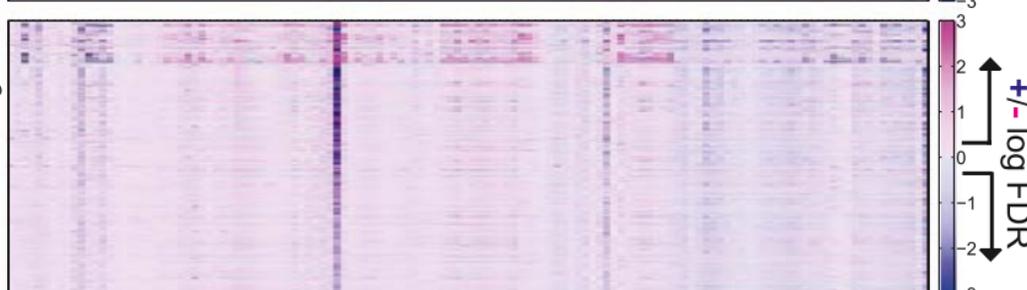
A**B**

A**B**

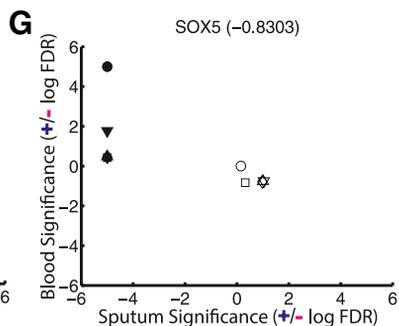
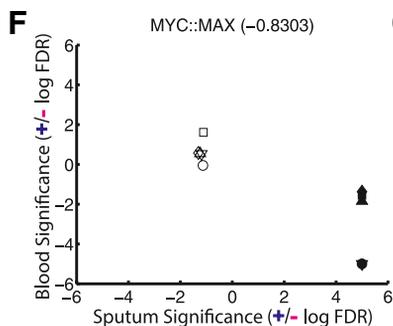
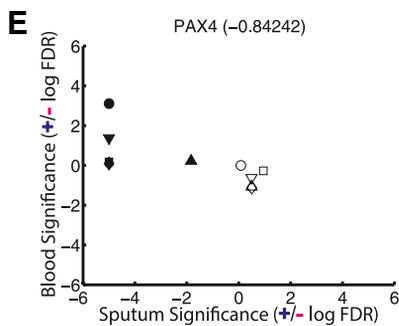
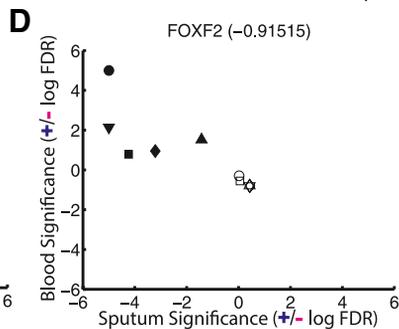
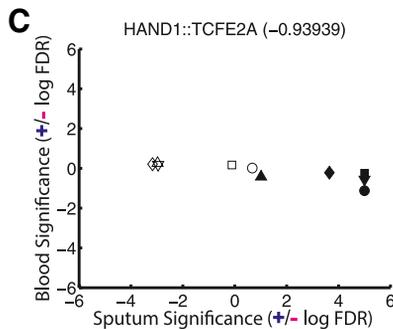
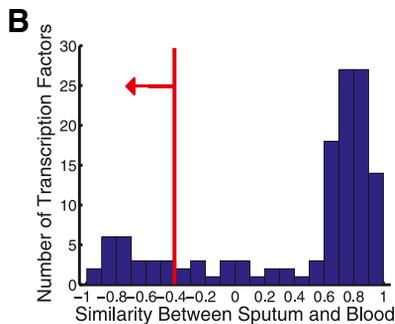
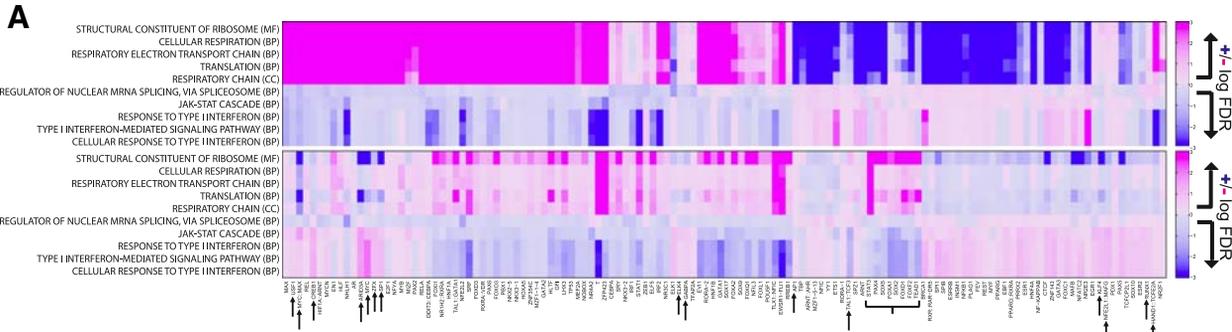
Sputum
(All Enriched Categories)

**C**

Blood
(All Enriched Categories)

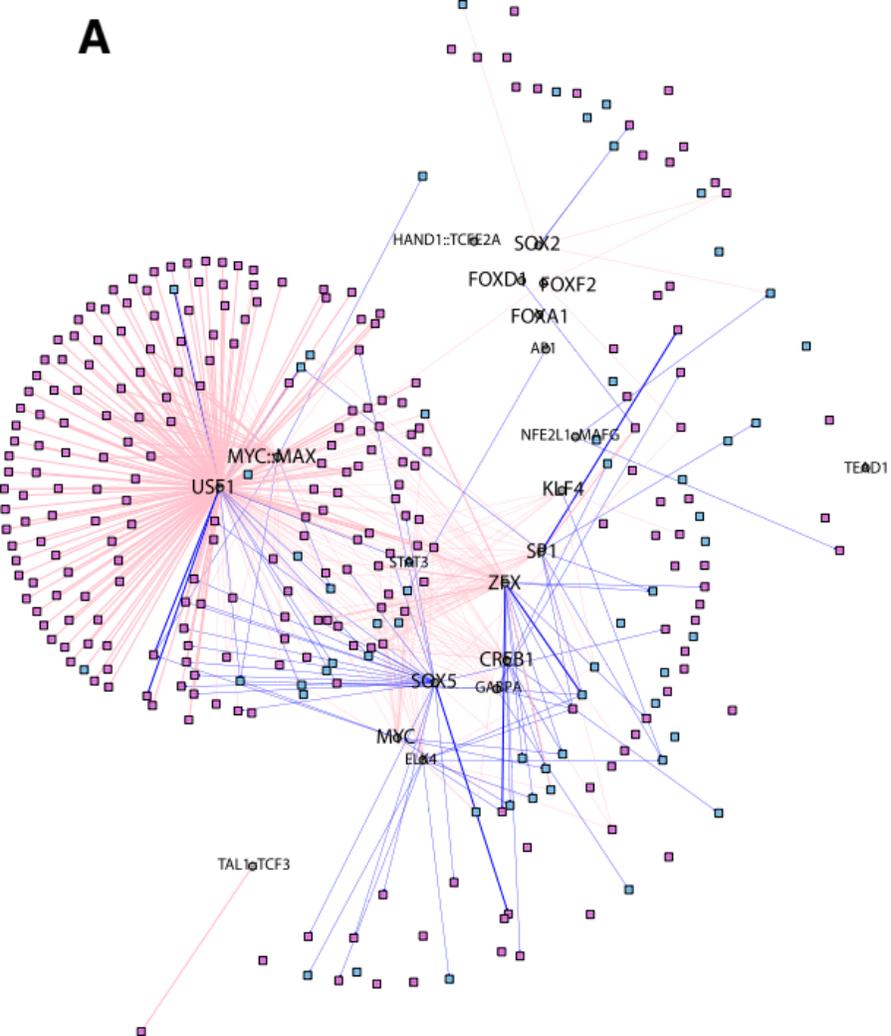
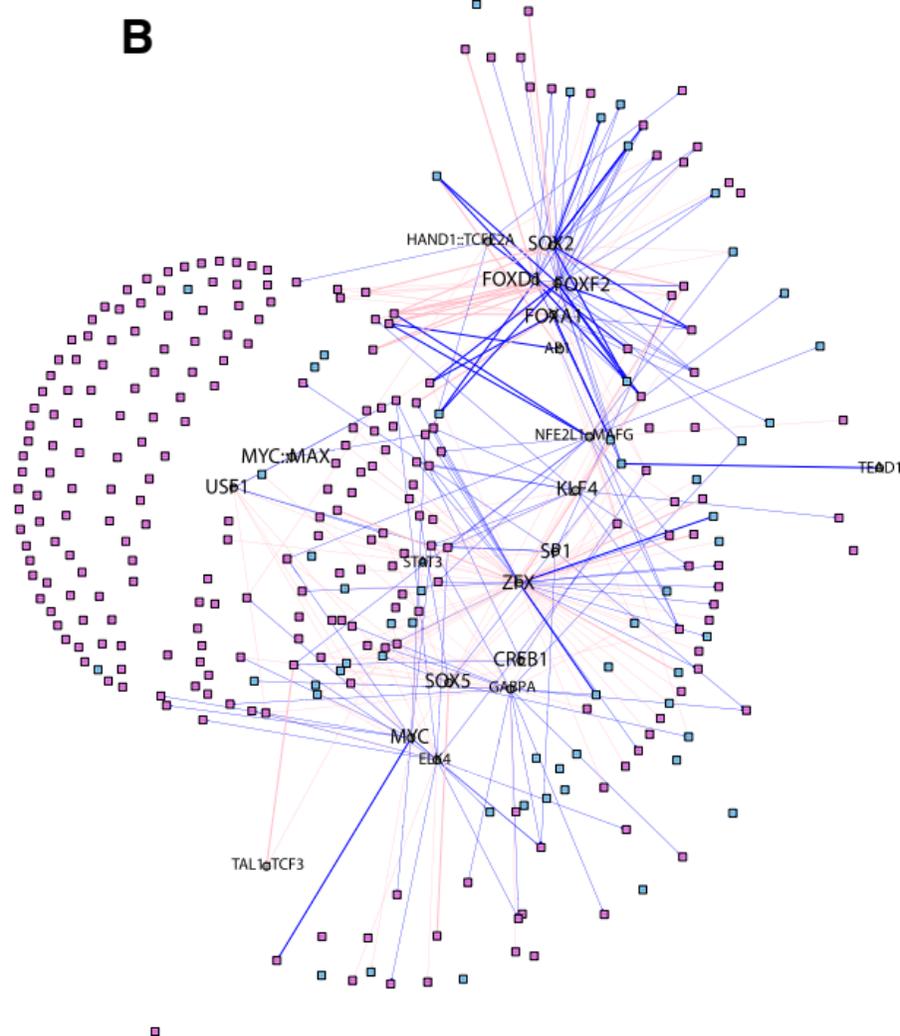


5001
 5002
 5003
 5004
 5005
 5006
 5007
 5008
 5009
 5010
 5011
 5012
 5013
 5014
 5015
 5016
 5017
 5018
 5019
 5020
 5021
 5022
 5023
 5024
 5025
 5026
 5027
 5028
 5029
 5030
 5031
 5032
 5033
 5034
 5035
 5036
 5037
 5038
 5039
 5040
 5041
 5042
 5043
 5044
 5045
 5046
 5047
 5048
 5049
 5050
 5051
 5052
 5053
 5054
 5055
 5056
 5057
 5058
 5059
 5060
 5061
 5062
 5063
 5064
 5065
 5066
 5067
 5068
 5069
 5070
 5071
 5072
 5073
 5074
 5075
 5076
 5077
 5078
 5079
 5080
 5081
 5082
 5083
 5084
 5085
 5086
 5087
 5088
 5089
 5090
 5091
 5092
 5093
 5094
 5095
 5096
 5097
 5098
 5099
 5100
 5101
 5102
 5103
 5104
 5105
 5106
 5107
 5108
 5109
 5110
 5111
 5112
 5113
 5114
 5115
 5116
 5117
 5118
 5119
 5120
 5121
 5122
 5123
 5124
 5125
 5126
 5127
 5128
 5129
 5130
 5131
 5132
 5133
 5134
 5135
 5136
 5137
 5138
 5139
 5140
 5141
 5142
 5143
 5144
 5145
 5146
 5147
 5148
 5149
 5150
 5151
 5152
 5153
 5154
 5155
 5156
 5157
 5158
 5159
 5160
 5161
 5162
 5163
 5164
 5165
 5166
 5167
 5168
 5169
 5170
 5171
 5172
 5173
 5174
 5175
 5176
 5177
 5178
 5179
 5180
 5181
 5182
 5183
 5184
 5185
 5186
 5187
 5188
 5189
 5190
 5191
 5192
 5193
 5194
 5195
 5196
 5197
 5198
 5199
 5200
 5201
 5202
 5203
 5204
 5205
 5206
 5207
 5208
 5209
 5210
 5211
 5212
 5213
 5214
 5215
 5216
 5217
 5218
 5219
 5220
 5221
 5222
 5223
 5224
 5225
 5226
 5227
 5228
 5229
 5230
 5231
 5232
 5233
 5234
 5235
 5236
 5237
 5238
 5239
 5240
 5241
 5242
 5243
 5244
 5245
 5246
 5247
 5248
 5249
 5250
 5251
 5252
 5253
 5254
 5255
 5256
 5257
 5258
 5259
 5260
 5261
 5262
 5263
 5264
 5265
 5266
 5267
 5268
 5269
 5270
 5271
 5272
 5273
 5274
 5275
 5276
 5277
 5278
 5279
 5280
 5281
 5282
 5283
 5284
 5285
 5286
 5287
 5288
 5289
 5290
 5291
 5292
 5293
 5294
 5295
 5296
 5297
 5298
 5299
 5300
 5301
 5302
 5303
 5304
 5305
 5306
 5307
 5308
 5309
 5310
 5311
 5312
 5313
 5314
 5315
 5316
 5317
 5318
 5319
 5320
 5321
 5322
 5323
 5324
 5325
 5326
 5327
 5328
 5329
 5330
 5331
 5332
 5333
 5334
 5335
 5336
 5337
 5338
 5339
 5340
 5341
 5342
 5343
 5344
 5345
 5346
 5347
 5348
 5349
 5350
 5351
 5352
 5353
 5354
 5355
 5356
 5357
 5358
 5359
 5360
 5361
 5362
 5363
 5364
 5365
 5366
 5367
 5368
 5369
 5370
 5371
 5372
 5373
 5374
 5375
 5376
 5377
 5378
 5379
 5380
 5381
 5382
 5383
 5384
 5385
 5386
 5387
 5388
 5389
 5390
 5391
 5392
 5393
 5394
 5395
 5396
 5397
 5398
 5399
 5400
 5401
 5402
 5403
 5404
 5405
 5406
 5407
 5408
 5409
 5410
 5411
 5412
 5413
 5414
 5415
 5416
 5417
 5418
 5419
 5420
 5421
 5422
 5423
 5424
 5425
 5426
 5427
 5428
 5429
 5430
 5431
 5432
 5433
 5434
 5435
 5436
 5437
 5438
 5439
 5440
 5441
 5442
 5443
 5444
 5445
 5446
 5447
 5448
 5449
 5450
 5451
 5452
 5453
 5454
 5455
 5456
 5457
 5458
 5459
 5460
 5461
 5462
 5463
 5464
 5465
 5466
 5467
 5468
 5469
 5470
 5471
 5472
 5473
 5474
 5475
 5476
 5477
 5478
 5479
 5480
 5481
 5482
 5483
 5484
 5485
 5486
 5487
 5488
 5489
 5490
 5491
 5492
 5493
 5494
 5495
 5496
 5497
 5498
 5499
 5500
 5501
 5502
 5503
 5504
 5505
 5506
 5507
 5508
 5509
 5510
 5511
 5512
 5513
 5514
 5515
 5516
 5517
 5518
 5519
 5520
 5521
 5522
 5523
 5524
 5525
 5526
 5527
 5528
 5529
 5530
 5531
 5532
 5533
 5534
 5535
 5536
 5537
 5538
 5539
 5540
 5541
 5542
 5543
 5544
 5545
 5546
 5547
 5548
 5549
 5550
 5551
 5552
 5553
 5554
 5555
 5556
 5557
 5558
 5559
 5560
 5561
 5562
 5563
 5564
 5565
 5566
 5567
 5568
 5569
 5570
 5571
 5572
 5573
 5574
 5575
 5576
 5577
 5578
 5579
 5580
 5581
 5582
 5583
 5584
 5585
 5586
 5587
 5588
 5589
 5590
 5591
 5592
 5593
 5594
 5595
 5596
 5597
 5598
 5599
 5600
 5601
 5602
 5603
 5604
 5605
 5606
 5607
 5608
 5609
 5610
 5611
 5612
 5613
 5614
 5615
 5616
 5617
 5618
 5619
 5620
 5621
 5622
 5623
 5624
 5625
 5626
 5627
 5628
 5629
 5630
 5631
 5632
 5633
 5634
 5635
 5636
 5637
 5638
 5639
 5640
 5641
 5642
 5643
 5644
 5645
 5646
 5647
 5648
 5649
 5650
 5651
 5652
 5653
 5654
 5655
 5656
 5657
 5658
 5659
 5660
 5661
 5662
 5663
 5664
 5665
 5666
 5667
 5668
 5669
 5670
 5671
 5672
 5673
 5674
 5675
 5676
 5677
 5678
 5679
 5680
 5681
 5682
 5683
 5684
 5685
 5686
 5687
 5688
 5689
 5690
 5691
 5692
 5693
 5694
 5695
 5696
 5697
 5698
 5699
 5700
 5701
 5702
 5703
 5704
 5705
 5706
 5707
 5708
 5709
 5710
 5711
 5712
 5713
 5714
 5715
 5716
 5717
 5718
 5719
 5720
 5721
 5722
 5723
 5724
 5725
 5726
 5727
 5728
 5729
 5730
 5731
 5732
 5733
 5734
 5735
 5736
 5737
 5738
 5739
 5740
 5741
 5742
 5743
 5744
 5745
 5746
 5747
 5748
 5749
 5750
 5751
 5752
 5753
 5754
 5755
 5756
 5757
 5758
 5759
 5760
 5761
 5762
 5763
 5764
 5765
 5766
 5767
 5768
 5769
 5770
 5771
 5772
 5773
 5774
 5775
 5776
 5777
 5778
 5779
 5780
 5781
 5782
 5783
 5784
 5785
 5786
 5787
 5788
 5789
 5790
 5791
 5792
 5793
 5794
 5795
 5796
 5797
 5798
 5799
 5800
 5801
 5802
 5803
 5804
 5805
 5806
 5807
 5808
 5809
 5810
 5811
 5812
 5813
 5814
 5815
 5816
 5817
 5818
 5819
 5820
 5821
 5822
 5823
 5824
 5825
 5826
 5827
 5828
 5829
 5830
 5831
 5832
 5833
 5834
 5835
 5836
 5837
 5838
 5839
 5840
 5841
 5842
 5843
 5844
 5845
 5846
 5847
 5848
 5849
 5850
 5851
 5852
 5853
 5854
 5855
 5856
 5857
 5858
 5859
 5860
 5861
 5862
 5863
 5864
 5865
 5866
 5867
 5868
 5869
 5870
 5871
 5872
 5873
 5874
 5875
 5876
 5877
 5878
 5879
 5880
 5881
 5882
 5883
 5884
 5885
 5886
 5887
 5888
 5889
 5890
 5891
 5892
 5893
 5894
 5895
 5896
 5897
 5898
 5899
 5900
 5901
 5902
 5903
 5904
 5905
 5906
 5907
 5908
 5909
 5910
 5911
 5912
 5913
 5914
 5915
 5916
 5917
 5918
 5919
 5920
 5921
 5922
 5923
 5924
 5925
 5926
 5927
 5928
 5929
 5930
 5931
 5932
 5933
 5934
 5935
 5936
 5937
 5938
 5939
 5940
 5941
 5942
 5943
 5944
 5945
 5946
 5947
 5948
 5949
 5950
 5951
 5952
 5953
 5954
 5955
 5956
 5957
 5958
 5959
 5960
 5961
 5962
 5963
 5964
 5965
 5966
 5967
 5968
 5969
 5970
 5971
 5972
 5973
 5974
 5975
 5976
 5977
 5978
 5979
 5980
 5981
 5982
 5983
 5984
 5985
 5986
 5987
 5988
 5989
 5990
 5991
 5992
 5993
 5994
 5995
 5996
 5997
 5998
 5999
 6000
 6001
 6002
 6003
 6004
 6005
 6006
 6007
 6008
 6009
 6010
 6011
 6012
 6013
 6014
 6015
 6016
 6017
 6018
 6019
 6020
 6021
 6022
 6023
 6024
 6025
 6026
 6027
 6028
 6029
 6030
 6031
 6032
 6033
 6034
 6035
 6036
 6037
 6038
 6039
 6040
 6041
 6042
 6043
 6044
 6045
 6046
 6047
 6048
 6049
 6050
 6051
 6052
 6053
 6054
 6055
 6056
 6057
 6058
 6059
 6060
 6061
 6062
 6063
 6064
 6065
 6066
 6067
 6068
 6069
 6070
 6071
 6072
 6073
 6074
 6075
 6076
 6077
 6078
 6079
 6080
 6081
 6082
 6083
 6084
 6085
 6086
 6087
 6088
 6089
 6090
 6091
 6092
 6093
 6094
 6095
 6096
 6097
 6098
 6099
 6100
 6101
 6102
 6103
 6104
 6105
 6106
 6107
 6108
 6109
 6110
 6111
 6112
 6113
 6114
 6115
 6116
 6117
 6118
 6119
 6120
 6121
 6122
 6123
 6124
 6125
 6126
 6127
 6128
 6129
 6130
 6131
 6132
 6133
 6134
 6135
 6136
 6137
 6138
 6139
 6140
 6141
 6142
 6143
 6144
 6145
 6146
 6147
 6148
 6149
 6150
 6151
 6152
 6153
 6154
 6155
 6156
 6157
 6158
 6159
 6160
 6161
 6162
 6163
 6164
 6165
 6166
 6167
 6168
 6169
 6170
 6171
 6172
 6173
 6174
 6175
 6176
 6177
 6178
 6179
 6180
 6181
 6182
 6183
 6184
 6185
 6186
 6187
 6188
 6189
 6190
 6191
 6192
 6193
 6194
 6195
 6196
 6197
 6198
 6199
 6200
 6201
 6202
 6203
 6204
 6205
 6206
 6207
 6208
 6209
 6210
 6211
 6212
 6213
 6214
 6215
 6216
 6217
 6218
 6219
 6220
 6221
 6222
 6223
 6224
 6225
 6226
 6227
 6228
 6229
 6230
 6231
 6232
 6233
 6234
 6235
 6236
 6237
 6238
 6239
 6240
 6241
 6242
 6243
 6244
 6245
 6246
 6247
 6248
 6249
 6250
 6251
 6252
 6253
 6254
 6255
 6256
 6257
 6258
 6259
 6260
 6261
 6262
 6263
 6264
 6265
 6266
 6267
 6268
 6269
 6270
 6271
 6272
 6273
 6274
 6275
 6276
 6277
 6278
 6279
 6280
 6281
 6282
 6283
 6284
 6285
 6286
 6287
 6288
 6289
 6290
 6291
 6292
 6293
 6294
 6295
 6296
 6297
 6298
 6299
 6300
 6301
 6302
 6303
 6304
 6305
 6306
 6307
 6308
 6309
 6310
 6311



Legend



A**B**

Additional files provided with this submission:

Additional file 1. Supplemental text and figures. This file includes PCA plots, the results of a gene expression and network-based analysis performed without removing sex chromosome genes and the results of several additional network analyses that complement those described in the main text (1204k)

<http://www.biomedcentral.com/content/supplementary/s12918-014-0118-y-s1.pdf>

Additional file 2. List of genes that have significantly different variance in their sputum expression profiles when comparing females and males (48k)

<http://www.biomedcentral.com/content/supplementary/s12918-014-0118-y-s2.txt>