

Article

Tissue-Specific Variations in Transcription Factors Elucidate Complex Immune System Regulation

Hengwei Lu, Yi-Ching Tang and Assaf Gottlieb * 

Center for precision health, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; hengwei.lu@uth.tmc.edu (H.L.); yi-ching.tang@uth.tmc.edu (Y.-C.T.)

* Correspondence: assaf.gottlieb@uth.tmc.edu

Abstract: Gene expression plays a key role in health and disease. Estimating the genetic components underlying gene expression can thus help understand disease etiology. Polygenic models termed “transcriptome imputation” are used to estimate the genetic component of gene expression, but these models typically consider only the *cis* regions of the gene. However, these *cis*-based models miss large variability in expression for multiple genes. Transcription factors (TFs) that regulate gene expression are natural candidates for looking for additional sources of the missing variability. We developed a hypothesis-driven approach to identify second-tier regulation by variability in TFs. Our approach tested two models representing possible mechanisms by which variations in TFs can affect gene expression: variability in the expression of the TF and genetic variants within the TF that may affect the binding affinity of the TF to the TF-binding site. We tested our TF models in whole blood and skeletal muscle tissues and identified TF variability that can partially explain missing gene expression for 1035 genes, 76% of which explains more than the *cis*-based models. While the discovered regulation patterns were tissue-specific, they were both enriched for immune system functionality, elucidating complex regulation patterns. Our hypothesis-driven approach is useful for identifying tissue-specific genetic regulation patterns involving variations in TF expression or binding.



Citation: Lu, H.; Tang, Y.-C.; Gottlieb, A. Tissue-Specific Variations in Transcription Factors Elucidate Complex Immune System Regulation. *Genes* **2022**, *13*, 929. <https://doi.org/10.3390/genes13050929>

Academic Editors: Rujiao Li, Yan Zhang and Jianzhong Su

Received: 22 March 2022

Accepted: 18 May 2022

Published: 23 May 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: transcriptional regulation; transcriptome imputation; transcription factor polymorphism

1. Introduction

Transcription regulation plays a critical role in cellular states, while transcriptional misregulation has been associated with a broad range of diseases [1]. Thus, the identification of genetic components that help explain gene transcription levels is valuable in order to generate accurate models that can further associate variations in gene expression with specific diseases. This is especially beneficial in tissues that are not readily accessible for direct measurement of gene expression.

Previously published methods for identifying gene regulation include expression quantitative trait loci (eQTLs) [2] and, more recently, transcriptome imputation techniques, such as PrediXcan or TWAS [3,4], which models the genetic component of observed gene expression using combinations of genetic locations. While eQTLs focus on single variants with strong effect size, transcriptome imputation techniques demonstrate that combinations of (possibly) smaller-effect genetic variants can explain transcription levels better.

Transcriptome imputation methods typically include only variants in *cis* with the gene they impute, owing in part to the large space of possible trans-acting locations. For many genes, however, these *cis* models explain only a small portion of the variability in gene expression. Indeed, a paper by Liu et al. [5] estimated that most heritability is driven by weak trans-eQTL single-nucleotide polymorphism (SNPs), supporting an approach to identify such sources of expression variability.

Despite the large space of trans-acting loci, a method by Wheeler et al. [6] used a data-driven approach to identify correlated trans-acting genes based on the PrediXcan

transcriptome imputation framework. They identified 55 significant associations between the expression of trans-genes and the imputed expression of 39 genes in blood tissue. However, their data-driven approach was limited to pairwise trans-gene correlations, required a strict false-discovery threshold rate due to the large hypothesis space, and, most importantly, the models lacked suggested mechanisms that limited the biological interpretation of the discovered correlations. Despite these limitations, Wheeler et al. found that trans-genes were enriched with transcription factor (TF) pathways, and target genes were enriched in transcription factor binding sites, supporting a hypothesis-driven approach that focuses on TFs as a source of expression variability.

Here, we develop a hypothesis-driven approach to identify associations between variations in TFs and variation in the transcription levels of their transcribed genes, suggesting a higher level of regulation. While regulation through polymorphism in TF-binding sites is well documented, as TFs regulate transcription by binding to *cis*-regulatory regions in the vicinity of the regulated genes [7], polymorphism in TFs themselves has been previously associated with diseases such as hypertension, coronary artery disease, type 2 diabetes mellitus, and dyslipidemia [8–10]. Thus, we hypothesize that variability in TFs may indirectly mediate an effect on disease phenotypes through their effect on the transcription levels of their transcribed genes. This approach addresses previous limitations in identifying trans-genes, since it provides a model that looks at the combinatorial effects of all the TFs affecting transcription of a gene instead of a single trans-gene, reduces the multiple hypotheses space, and provides a mechanistic interpretation of these trans-models.

We considered two possible mechanisms where variability in a TF can lead to variability in the expression of a gene: one mechanism considers variations in the expression of TFs to correlate with variations in the expression of their transcribed genes. Another mechanism suggests that non-synonymous single-nucleotide polymorphisms (SNPs) within TFs with predicted deleterious effects may affect the binding affinity and thus the transcription levels of the transcribed genes. We thus developed two distinct models to capture these potential mechanisms. With these models, we are able to account for transcriptional variability that is only partially explained by *cis* imputation models and suggest potential biological mechanisms for this higher-level regulation. We applied our models to two tissues, skeletal muscle and whole blood, identifying 1035 hit genes where the TF model can explain a significant portion of the variability beyond the *cis* models, with 76% explaining more variability than the *cis* models alone.

To summarize, our hypothesis-driven approach introduces three contributions: (1) we detect cumulative weak effects that individual trans-eQTLs may have insufficient power to detect; (2) our hypothesis-driven approach is biologically interpretable; and (3) it reduces the number of hypotheses relative to data-driven approaches. Thus, our models enable the identification of higher-level regularization of TFs, which may further suggest clinical implications through their effect on disease-associated genes.

2. Materials and Methods

2.1. Data

Genotype and expression data from the Genotype-Tissue Expression Project (GTEx) version 8 [11] were retrieved from dbGaP. Transcription factors and their transcribed genes (204,999 unique gene–TF pairs, see Table 1 for statistics) were assembled from three sources: Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining (TR-RUST V2) [12], the Human Transcriptional Regulation Interaction Database (HTRIdb) [13], and the Regulatory Network Repository of Transcription Factor and microRNA Mediated Gene Regulations (RegNetwork) [14]. Genomic positions of the TFs were computed based on the Human Genome Assembly version 37 (GRCh37). Functional annotations for non-synonymous SNPs were retrieved from SnpEff v4.3 [15].

Table 1. Summary statistics of input data per tissue: transcription factors and associated SNPs.

	Skeletal Muscle	Whole Blood
Number of samples	706	670
Number of expressed genes	21,031	20,315
Number of expressed genes with an associated TF (% of genes)	11,130 (53%)	10,563 (52%)
Number TFs per gene	10.8 ± 8.4	10.9 ± 8.7
Number of nsSNP* per TF	1.55 ± 1.77	1.57 ± 1.99

* nsSNP: non-synonymous SNP.

2.2. Constructing TF Models

The constructed TF models follow a five-step protocol (Figure 1).

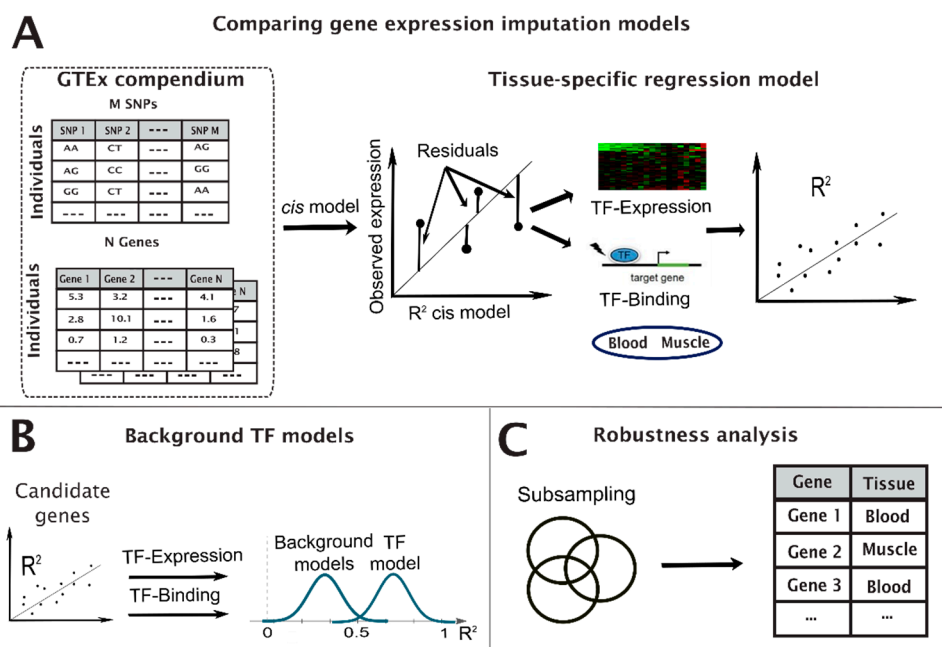


Figure 1. Illustration of the filtering pipeline to identify TF hit genes. We computed the baseline *cis* model and compared them with one of the three TF models to identify candidate genes per TF model (A), computed a background model for each of the candidate genes to test their significance (B), and conducted a robustness test to validate the results (C).

Step 1, normalized observed genes expression: We followed the PrediXcan recommended normalization to retain only the genetic component of the observed expression by accounting for gender, sequencing platform, the top 3 principal components of the genotype dosages, and the 15 probabilistic estimation of expression residuals (PEER) factors [16] for each gene's observed expression. The residuals between the linear model formed by these factors and the observed expression values from GTEx are the normalized gene expression considered in the following steps.

Step 2, apply PrediXcan: We applied the pre-trained *cis* models (baseline) developed in the PrediXcan method [16], coded in Python and R. PrediXcan methods predict the expression values of each gene in the selected tissue based on their genotypes.

Step 3, build a model for candidate genes: We modeled the residuals between the normalized observed expression and the PrediXcan model of the genes using regularized LASSO regression [17]. The independent variables for the TF-expression model were the expression of the TFs that are associated with each candidate gene and, for the TF-binding

model, the dosages of a set of non-synonymous SNPs within these TFs. In order to focus on the SNPs most likely to affect TF binding, we retained only the non-synonymous SNPs that are considered deleterious according to SIFT [18] (SIFT score < 0.05).

Step 4, background models: We filtered the candidate genes based on significance in two background models: The first background model shuffled the residuals and re-trained the models one hundred times. For the second filter, we evaluated the significance of each candidate gene relative to the random selection of non-associated TFs. We selected 100 random sets of TFs for each candidate gene (each set with the same number of TFs as the true set of TFs associated with the gene) and computed the R^2 for each random set of TFs. We retained only genes that passed both background models with significance level of 0.05 (adjusted p -value using Benjamini Hochberg false-discovery rate [19]).

Step 5, robustness check: A robustness check was conducted by randomly removing 10% of the samples and re-running the pipeline. Genes that were discovered in more than 50% of ten robustness runs (i.e., at least five of the runs) were declared hit genes.

2.3. Statistical Analysis

Statistical significance of overlap with other works was computed using the hypergeometric test. Enrichment of hit genes with biological functions was performed based on the hypergeometric test using the ToppGene suit [20].

3. Results

3.1. Identifying Gene Expression Associated with TF Variability

We hypothesized that variability in TFs may be associated with variability in the expression of their transcribed genes. We considered two potential mechanisms to test this hypothesis: (1) variability in transcription levels of the TF that subsequently translate to variability in the transcription levels of the transcribed gene (Figure 2A); or (2) SNPs within TFs can affect change in binding affinity of the TF to the *cis*-binding sites of their transcribed genes (Figure 2B). We refer to these models as TF-expression and TF-binding, respectively (Section 2, Figure 1).

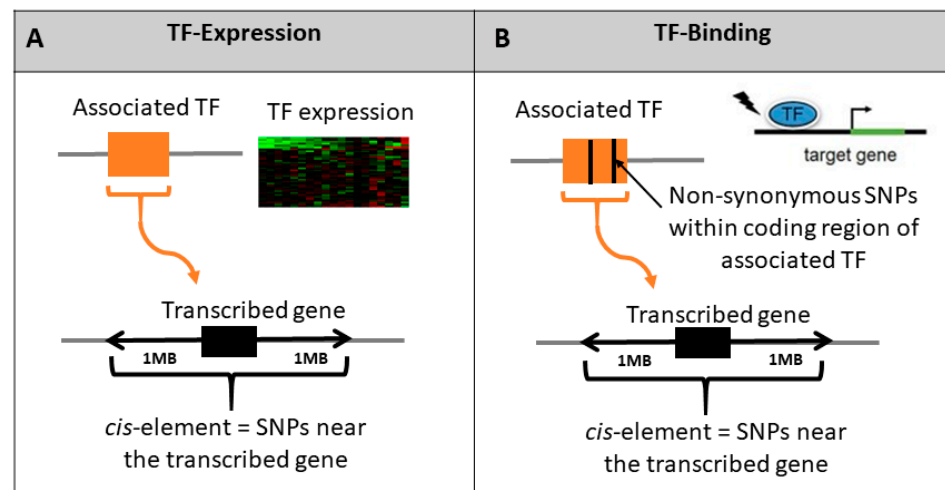


Figure 2. Illustration of the three tested hypotheses regarding the effect of TF genetic polymorphism on the expression of their transcribed genes. We test whether polymorphism in the TF (orange box) affects the transcription levels of the transcribed gene (black box) by testing the added effect of each TF model relative to the baseline *cis* model. TF-expression model includes association of the TF expression with imputed gene expression (A); TF-binding model includes non-synonymous SNPs within the associated TF boundary (B).

We developed a pipeline that tests model significance and robustness with two background models and a robustness test to reduce false positives (Section 2). To demonstrate

our method, we focused on the two tissues with the largest sample size in the genome tissue expression (GTEx) compendium: skeletal muscle and whole blood. Table 1 provides sample sizes and statistics for the two tissues.

We term “hit genes” as genes where the model (TF-expression or TF-binding) provided a significant improvement in explaining the gene transcription levels over the base *cis* model, passing all significance tests (Section 2, Figure 1). Following these tests, we identified 1095 hit genes (1035 unique), with 379 and 419 hit genes from the TF-expression model in muscle and blood tissues, respectively, and 155 and 138 hit genes from the TF-binding model on muscle and whole-blood tissues, respectively (Tables S1 and S2). Forty-one of the hit genes (4%) were discovered in both tissues by the TF-expression model, while none were shared between tissue in the TF-binding models, suggesting it is more tissue-specific.

Notably, the TF-expression models were more robust to the robustness test, with only 6 genes in skeletal muscle and 17 in blood that did not pass this filter, while in the TF-binding models, 183 genes in muscle and 167 genes in blood that did not pass the robustness test (see Section 2). Only one gene from muscle shares SNPs used in the *cis* model with the TF SNPs used in our TF-binding model, and none are shared in blood.

There was a low correlation between the *cis*-based R^2 and R^2 of the TF-expression models (Pearson ρ ranged between -0.12 and 0.06 , with significant p -values after correction for false-discovery rate only for blood TF-expression, see also Figure 3), suggesting that the *cis* and trans effects are generally independent.

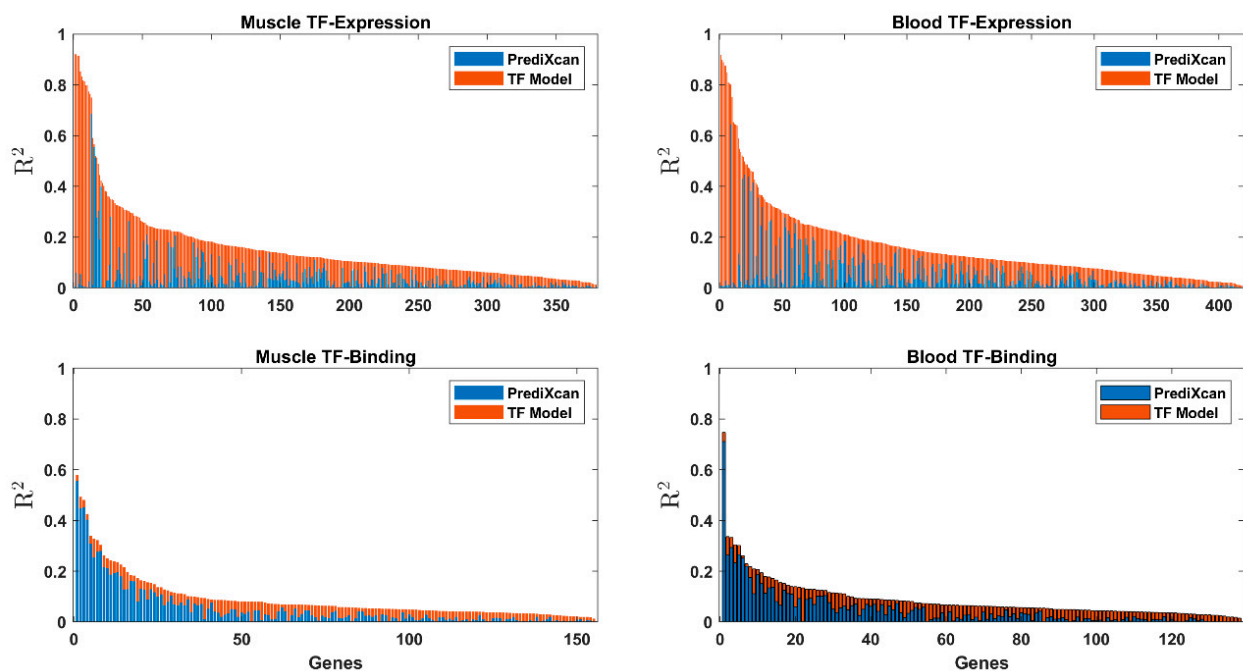


Figure 3. Stacked bar graph of the R^2 results of the *cis* baseline model (blue) and the skeletal muscle and whole-blood TF-expression and TF-binding (orange).

3.2. Comparison with Other Methods

We first compared the large-scale trans-eQTLs dataset of Vösa et al. [21]. They identified 3966 genes with trans-eQTLs in whole blood, 2909 of which were tested in our whole-blood set. While they reported very low reproducibility in GTEx (between 0.07% and 0.09%), we still found that 175 (of the 419 hit genes, 42%) discovered in whole-blood TF-expression also had trans-eQTLs discovered in their whole-blood sample (hypergeometric p -value $< 10^{-10}$). Additionally, 30 out of 138 genes (22%) were discovered through the TF-binding model in blood, but this overlap is not statistically significant. While the eQTLs of Vösa et al. [21] were discovered in whole-blood tissue, we still found significant overlap with genes discovered with our TF-expression model within muscle tissue

(128/379, $p < 3 \times 10^{-6}$). This overlap was not significant in muscle TF-binding (45/155 shared genes).

One gene out of seven from the dataset of Vösa et al. [21] that were replicated in GTEx whole-blood tissue, hemoglobin subunit γ 1 or γ -globin (HBG), overlapped with our hit genes in the blood TF-expression model, but the overlap was not statistically significant. Similarly, one gene, ARL6IP1, was replicated in GTEx skeletal muscle and also detected by our skeletal muscle TF-expression model (hypergeometric test $p < 0.03$). Both of these genes had a larger trans effect ($R^2 = 0.04$ and 0.13 in blood and muscle, respectively) than the *cis* model ($R^2 = 0.008$ and 0.08). We highlight these two genes below.

The highest-weighted TF in the TF-expression model of whole blood for HBG1 is Krüppel-like factor 3 (KLF3) (weight = -0.13). KLF3 represses a subset of KLF1/EKLF target genes and is required for proper erythroid (cells differentiated from hematopoietic stem cells) maturation [22]. KLF1 in human and mouse adult erythroid progenitors markedly increases human γ -globin expression ratios [23], while EKLF protein activates γ -globin in the fetal-like erythroleukemia cell line [24]. Interestingly, previous evidence shows that HBG1, KLF1, and KLF3 are up-regulated in the β globin knockout human erythroid progenitor cell model, relative to controls [25] and in human myeloid leukemia cells [26]. Finally, other TFs in the model are NF-YA (weight = 0.11), GATA-2 (weight = -0.08), and GATA-1 (weight = 0.04), where NF-Y recruits both the transcription activator and repressor to modulate the tissue-specific expression of the human γ -globin gene, with GATA-2 part of the transcription activator hub and GATA-1 as part of the transcription repressor hub [27–29].

Mutations in ARL6IP1 are associated with spastic paraplegia 61, a rare inherited disorder that causes weakness and stiffness in the leg muscles [30,31]. The TFs with the highest weight in the model for ARL6IP1 are GATA-2, GATA-3, and PPARG (weights = 0.11 , 0.13 , and 0.14 , respectively), all with their reported association with skeletal muscle tissue. Specifically, GATA-2 and GATA-3 are expressed in skeletal muscle cells [32], where GATA-2 is a unique marker of hypertrophy in skeletal muscle fibers [33,34], and GATA-3 promotes the invasion of tumors into skeletal muscles [35]. Additionally, PPARG inhibits NF κ B-dependent transcriptional activation in skeletal muscle [36], and PPARG polymorphism is associated with improved glucose utilization in skeletal muscles [37]. Interestingly, the trans-eQTLs associated with ARL6IP1 (rs6592965 and rs12718598) both reside in the intron of the TF gene IKZF1. While IKZF1 does not bind directly to ARL6IP1, GATA-3 is up-regulated in IKZF1-deleted samples [38], suggesting potential co-regulation between the TFs.

Our second comparison was with the results of Wheeler et al. [6], who used a data-driven approach to test the association of the expression of trans-acting genes with the imputed expression of other genes (via PrediXcan). They identified 39 genes with trans associations (22 and 33 of which are part of our initial gene set of muscle and blood tissues, respectively). Seven of these genes were predicted also by our pipeline, with five of these shared genes identified in whole blood (CRISP3, NBL1, ASTL, and MANSC1 in blood TF-expression and ZNF135 in blood TF-binding; hypergeometric probability, $p < 0.03$) and two, CLEC12A and RNASE6, in muscle tissue TF-expression (but not statistically significant). This corresponds well with our expectations, as Wheeler's genes were discovered in whole-blood tissue.

Out of these seven genes shared with Wheeler et al., we highlight the two genes with the highest model R^2 in blood (cysteine-rich secreted protein 3 (CRISP3)) and ribonuclease A family member K6 (RNASE6) in muscle tissue. CRISP3 is activated in mouse B cells and may be a potential biomarker of multiple myeloma [39,40]. The TF with the largest weight in the model is the ETS-related gene (ERG), which is shown to be essential for early B lymphoid differentiation [41] and nuclear expression of ERG in a vast majority of myeloma cells [42]. Furthermore, increased expression of ERG is indicative of poor prognosis of acute lymphoblastic leukemia and cytogenetically normal acute myeloid leukemia [43].

The second gene shared with Wheeler et al. is RNASE6, showing a substantial increase in explained variability ($R^2 = 0.18$, while PrediXcan $R^2 = 0.02$) in muscle tissue. Spi-1 proto-oncogene (SPI-1) is the TF with the highest weight in the TF-expression model of RNASE6. The detected genes and their TF are associated with skeletal muscle [44–46].

3.3. Hit Genes in Expression Models Are Enriched with Immune Response

The TF-expression models were highly enriched in immune system Gene Ontology biological processes, including top biological processes of activation of immune cells, including leukocyte, myeloid cells, neutrophil, and granulocyte activation involved in immune response in muscle tissue (Benjamini–Hochberg (B&H) FDR-adjusted q -value $< 3 \times 10^{-17}$) and immune system development, regulation of the immune system, and lymphocyte and leukocyte proliferation in blood tissue (B&H FDR-corrected q -value $< 8 \times 10^{-7}$), as well as blood-specific terms such as hemopoiesis (q -value $< 2 \times 10^{-7}$) and a human phenotype of hemolytic anemia (q -value $< 2 \times 10^{-3}$) (Table S2). Additionally, genes discovered in both TF-expression model tissues are enriched for the cellular location of the major histocompatibility complex (MHC) protein complex (B&H FDR-adjusted q -value $< 3 \times 10^{-6}$ and q -value $< 2 \times 10^{-3}$ for muscle and blood, respectively). MHC class II is known to be expressed in skeletal muscle [47,48] and is tightly regulated by a variety of transcription factors [49,50].

The TF-expression model in both tissues was also enriched for the autoimmune disease systemic lupus erythematosus (SLE) (with 76 and 65 associated genes and FDR-adjusted q -value $< 6 \times 10^{-15}$ and $p < 3 \times 10^{-9}$ for muscle and blood tissues, respectively). Indeed, SLE patients have a high prevalence of leukopenia, lymphopenia, and neutropenia [51], as well as inflammatory myositis [52]. Additionally, greater skeletal muscle mitochondrial dysfunction is found among fatigued patients with SLE [53].

In contrast to the TF-expression model, the TF-binding model in both tissues has no enriched Gene Ontology biological processes. TF-Binding in muscle is enriched for coated vesicle membrane and for the cellular compartment of ER to the Golgi transport vesicle membrane (q -value < 0.02). Indeed, extracellular vesicles have a role in skeletal muscle [54,55]. During skeletal muscle differentiation, the Golgi complex undergoes dramatic reorganization [56], and muscle impulse activity plays a major role in regulating the organization of the Golgi complex in the extra-junctional region of muscle fibers [57]. We highlight two Golgi-related hit genes in muscle TF-binding. The first is Bet1 Golgi vesicular membrane trafficking protein (BET1), where low BET1 protein levels are associated with impaired ER-to-Golgi transport in congenital muscular dystrophy [58]. The second example is NOTCH2, where Notch signaling plays a critical role in the regulation of embryonic and post-natal skeletal myogenesis, and one of its activation stages is mediated in the Golgi apparatus [59].

As evident from the low overlap between tissues and models, the discovered genes were tissue- and mechanism-specific, suggesting fine-tuned regulation (Tables S1 and S3). The exceptions are four hit genes (MORN1, CYP11B1, FAM222A, and FAM91A1), discovered by both TF models in blood tissue, and five genes (HLA-DRB1, NCF2, SNX20, SIAE, and BCHE), discovered in muscle tissue by both models.

We highlight the major histocompatibility complex, class II, DR β 1 (HLA-DRB1), which was discovered in both tissues in muscle tissue by both models. Our model explains the same amount of variability of HLA-DRB1 as the *cis* model in both tissues. PrediXcan obtained $R^2 = 0.16$ and $R^2 = 0.11$ on blood and muscle tissues, respectively, and our TF-expression model obtained the same in each tissue on top of the PrediXcan model. The TF-binding model obtained a smaller R^2 of 0.02 in the muscle tissue (Table S1, Figure 3). Indeed, it was previously reported that both *cis*- and trans-regulatory polymorphisms modulate the expression of HLA-DRB1 [60,61].

The connection between variations in HLA-DRB1 and its TFs is further supported by reported associations between the expression of HLA-DRB1 and the expression of its TFs with the same disease. Specifically, the TF with the highest weight in the TF-

expression model ($\beta = 0.23$ in muscle and $\beta = 0.18$ in blood) is CIITA, where variability in the CIITA gene has been reported to interact with specific mutations of HLA-DRB1 to increase the risk for multiple sclerosis [62,63] and type 1 diabetes [64,65]. Additionally, the expression of HLA-DRB1 is also predictive of type 2 diabetes [66], as is the expression of CIITA and another HLA-DRB1 TF, RFX5, which determine susceptibility to type 2 diabetes mellitus [67]. In relation to the specificity of the tissues, skeletal muscle insulin resistance is the primary defect in type 2 diabetes [68,69]. While the TF-binding model explains lower variability in HLA-DRB1 than TF-expression, it is interesting to note that vitamin d receptor (VDR) gains high weight in the TF-binding model (with rs141329158 as the non-synonymous SNP), with reported interaction with HLA DRB1 in type 1 diabetes patients from North India [70].

3.4. Explanatory Trans Associations Display Tissue-Specific Regulation of Immune Response

We define explanatory trans associations as hit genes where our TF models explain at least the same portion of the gene expression (in terms of R^2) as the cis-based model. There were 787 hit genes (76%) with explanatory trans associations (Table S1). In total, 36 (3.5%) of them were discovered in both tissues, and an additional 9 hit genes are explanatory trans associations in one tissue but not in another.

The hit genes with the highest explanatory trans associations are transporter 1, ATP binding cassette subfamily B member (TAP1), with $R^2 = 0.63$ in whole blood and $R^2 = 0.31$ in skeletal muscle (PrediXcan $R^2 = 0.02$ in both tissues). Regulation of the TAP1 gene is critical for the initiation and continuation of a cellular immune response [71]. Furthermore, STAT1 is necessary for the activation of TAP1 gene expression and the initiation of cellular immune responses [71], where STAT1 is the TF with the highest weight in our TF model in both tissues (weights = 0.33 and 0.29 in blood and muscle, respectively). Another TF with a high weight in both tissues is IRF1 (weights of 0.16 and 0.22 in blood and muscle, respectively), where regulation of TAP1 by IRF1 explains the deficiency of CD8+ t cells in IRF1 deficient mice [72].

Another top hit gene with the highest explanatory trans associations in whole blood is interferon regulatory factor 9 (IRF9) ($R^2 = 0.63$, where PrediXcan R^2 is only 0.01). The TFs with the highest weights in the IRF9 model are STAT1 ($\beta = 0.28$) and STAT2 ($\beta = 0.24$). Supporting their large weight in the model, the signaling cascade activated by type I and type III interferons is dominated by the formation of a complex comprising IRF9, STAT1, and STAT2 [73–75]. Type I and III interferons are known to promote anti-viral immune responses [76], with type III also specifically with the blood–brain barrier [77] and regulator of neutrophil function [78]. While IRF9 is integral in mediating the type I interferon anti-viral response and the role of IRF9 in many important non-communicable diseases has just begun to emerge, the regulation of IRF9 during these conditions is not well understood [79]. What is known is that high levels of IRF9 and STAT1/STAT2 drive a prolonged response of the initial anti-viral response and also provide resistance to DNA damage [80,81].

IRF9 and STAT1 play a role in the pathogenesis of systemic lupus erythematosus [82,83], which was previously mentioned to be highly enriched with the blood TF-expression model hit genes. The second top hit gene with a major trans effect in whole blood is interferon-induced protein with tetratricopeptide repeats 3 (IFIT3) ($R^2 = 0.47$) with major trans effects but lower variability explained in skeletal muscle ($R^2 = 0.11$). Abnormal elevations in IFIT3 are associated with stimulants of interferon genes signaling in human systemic lupus erythematosus monocytes [84,85]. While the highest-weighted TF in the IFIT3 model is IRF9 in both tissues (weights = 0.48 and 0.16, in blood and muscle tissues), STAT1 and STAT2 have much higher weights in blood (weights = 0.17 for both) in blood but negligible ones in muscle, suggesting a tissue-specific regulation of IFIT3.

4. Discussion

We hypothesized that variability in TFs may be associated with variability of transcription levels in their transcribed genes. We tested our hypothesis using two types of TF models, based on proposed mechanisms for the effect—a model assuming that change in the expression levels of the TF affects the expression levels of the gene, and a model considering the effect of change in binding affinity between TF and TF-binding site.

Our focus here was on describing the methodology, and we thus demonstrated it on the selected set of tissues. Sample size can affect the performance of gene imputation [86]; thus, we selected the tissues with the largest sample size in GTEx, enabling us to test the robustness of the results by repeating the analysis on sub-samples of the data. We further passed our models through rigorous background model tests to reduce type I errors.

Our work suggests that the discovered hit genes are tissue specific, which corresponds well to the previous observation of Lopes-Ramos et al. [87] that TF regulation is tissue specific, including observations in the whole-blood tissue, used in our study. Additionally, there are different genes discovered in the TF-expression and the TF-binding models, suggesting that these mechanisms may be complementary.

Based on the TF-binding results, polymorphism within TFs affects a smaller fraction of genes than with TF-expression models. This result may be explained by the complexity of the regulatory plan. As TFs typically transcribe several genes, major changes in TF binding or regulation may result in significant effects on several genes and cellular processes, which could be detrimental to the individual. Such genetic variation may suffer from evolutionary pressure to attenuate its effect. Given the enrichment of immune system hits in our TF-expression models, we could speculate that in the case of the immune system, a complex regulation plan that is not “hard-coded” into the genome but rather regulated on the transcription level might provide more benefits. This immune-specific regulation complexity was previously discussed, but more research is required to quantify it [88,89].

The lower number of hit genes identified by the TF-binding model than hit genes identified by the TF-expression model could potentially be a result of our focus on deleterious non-synonymous SNPs. There are two reasons why we focused on these SNPs. The first reason concerns the biological interpretability of our model. As we built this model under the assumption that effects on gene expression are through altered binding affinity, variants predicted to be deleterious would be the initial suspects, as they are expected to have the largest effect on the TF conformation and ultimately the binding affinity. The second reason is that incorporating a more focused set of variants helps in avoiding overfitting the models in the high-dimensional sparse space of variants. Thus, our detected hit genes from the TF-binding model could be regarded as an underestimate of the *trans* effect of variants within TFs.

While we propose a hypothesis-driven approach, our findings only establish associations and not causation. While further validation is required to establish causal relationships, our findings can be useful in prioritizing genes and TFs for experimental validation.

5. Conclusions

Understanding how variations in TFs regulate gene expression can offer key insights into the transcriptional regulation plan. Here, we demonstrated how we identify expression variations of TFs and combinations of SNPs within TFs that associate with gene expression levels, some with potential associations with diseases. Our models can help in two ways. The first way is by introducing new factors that can be associated with human phenotypes in large association studies, potentially improving phenotype prediction models. The second way is by generating new hypotheses for genetic architecture that can be further explored in additional experiments. Thus, our models should be integrated into future gene expression imputation methods, allowing us to improve association studies and gain a better understanding of human genetic architecture.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13050929/s1>, Table S1: Hit genes and their corresponding R² performance for the skeletal muscle and whole-blood TF-expression and TF-binding models. Table S2: Hit genes and their corresponding TFs with their LASSO model weights for the skeletal muscle and whole blood TF-expression and TF-binding models. Table S3: Model weights for TFs in TF-Expression model and nsSNPs for the TF-Binding model for all the hit genes.

Author Contributions: Conceptualization, A.G.; Methodology, A.G.; Software, H.L., Y.-C.T.; Validation, H.L., A.G.; Formal Analysis, H.L.; Investigation, H.L.; Resources, A.G.; Data Curation, H.L.; Writing—Original Draft Preparation, A.G.; Writing—Review and Editing, A.G., H.L. and Y.-C.T.; Visualization, A.G.; Supervision, A.G.; Project Administration, A.G.; Funding Acquisition, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the University of Texas Health Science Center at Houston (Protocol Code: HSC-SBMI-19-0872; approval date: 21 October 2019).

Informed Consent Statement: Not applicable.

Data Availability Statement: GTEx data are available from dbGaP (Accession Number: phs000424.v7.p2). Transcription factors and their transcribed genes are publicly available from Transcriptional Regulatory Relationships Unraveled by Sentence-based Text Mining (TRRUST V2) [12], the Human Transcriptional Regulation Interaction Database (HTRI) [13], and the Regulatory Network Repository of Transcription Factor and microRNA Mediated Gene Regulations (RegNetwork) [14]. Code is freely available through GitHub: [90].

Acknowledgments: We would like to thank Arif Harmanci for constructive discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251. [[CrossRef](#)] [[PubMed](#)]
2. Nica, A.C.; Dermitzakis, E.T. Expression quantitative trait loci: Present and future. *Phil. Trans. R. Soc. B* **2013**, *368*, 20120362. [[CrossRef](#)] [[PubMed](#)]
3. Gusev, A.; Ko, A.; Shi, H.; Bhatia, G.; Chung, W.; Penninx, B.W.; Jansen, R.; De Geus, E.J.; Boomsma, D.I.; Wright, F.A. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **2016**, *48*, 245. [[CrossRef](#)] [[PubMed](#)]
4. Gamazon, E.R.; Wheeler, H.E.; Shah, K.P.; Mozaffari, S.V.; Aquino-Michaels, K.; Carroll, R.J.; Eyler, A.E.; Denny, J.C.; Nicolae, D.L.; Cox, N.J. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **2015**, *47*, 1091–1098. [[CrossRef](#)]
5. Liu, X.; Li, Y.I.; Pritchard, J.K. Trans effects on gene expression can drive omnigenic inheritance. *Cell* **2019**, *177*, 1022–1034.e1026. [[CrossRef](#)]
6. Wheeler, H.E.; Ploch, S.; Barbeira, A.N.; Bonazzola, R.; Andaleon, A.; Fotuhi Siahpirani, A.; Saha, A.; Battle, A.; Roy, S.; Im, H.K. Imputed gene associations identify replicable trans-acting genes enriched in transcription pathways and complex traits. *Genet. Epidemiol.* **2019**, *43*, 596–608. [[CrossRef](#)]
7. Hobert, O. Gene regulation by transcription factors and microRNAs. *Science* **2008**, *319*, 1785–1786. [[CrossRef](#)]
8. Fujimaki, T.; Oguri, M.; Horibe, H.; Kato, K.; Matsuoka, R.; Abe, S.; Tokoro, F.; Arai, M.; Noda, T.; Watanabe, S. Association of a transcription factor 21 gene polymorphism with hypertension. *Biomed. Rep.* **2015**, *3*, 118–122. [[CrossRef](#)]
9. Palizban, A.; Rezaei, M.; Khanahmad, H.; Fazilati, M. Transcription factor 7-like 2 polymorphism and context-specific risk of metabolic syndrome, type 2 diabetes, and dyslipidemia. *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.* **2017**, *22*, 40. [[CrossRef](#)]
10. Hamed, W.A.; Hammouda, G.E.; El-Hefnawy, S.M. Transcription factor 21 gene polymorphism in patients with coronary artery disease. *Res. Rep. Clin. Cardiol.* **2016**, *55*, 13–18. [[CrossRef](#)]
11. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580. [[CrossRef](#)] [[PubMed](#)]
12. Han, H.; Cho, J.-W.; Lee, S.; Yun, A.; Kim, H.; Bae, D.; Yang, S.; Kim, C.Y.; Lee, M.; Kim, E. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **2017**, *46*, D380–D386. [[CrossRef](#)] [[PubMed](#)]
13. Bovolenta, L.A.; Acencio, M.L.; Lemke, N. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genom.* **2012**, *13*, 405. [[CrossRef](#)] [[PubMed](#)]

14. Liu, Z.-P.; Wu, C.; Miao, H.; Wu, H. RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* **2015**, *2015*, bav095. [[CrossRef](#)] [[PubMed](#)]
15. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)]
16. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]
17. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
18. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [[CrossRef](#)]
19. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [[CrossRef](#)]
20. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **2009**, *37*, W305–W311. [[CrossRef](#)]
21. Vösa, U.; Claringbould, A.; Westra, H.-J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Brugge, H. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **2021**, *53*, 1300–1310. [[CrossRef](#)] [[PubMed](#)]
22. Funnell, A.P.; Norton, L.J.; Mak, K.S.; Burdach, J.; Artuz, C.M.; Twine, N.A.; Wilkins, M.R.; Power, C.A.; Hung, T.-T.; Perdomo, J. The CACCC-binding protein KLF3/BKLF represses a subset of KLF1/EKLF target genes and is required for proper erythroid maturation in vivo. *Mol. Cell. Biol.* **2012**, *32*, 3281–3292. [[CrossRef](#)] [[PubMed](#)]
23. Zhou, D.; Liu, K.; Sun, C.-W.; Pawlik, K.M.; Townes, T.M. KLF1 regulates BCL11A expression and γ -to β -globin gene switching. *Nat. Genet.* **2010**, *42*, 742–744. [[CrossRef](#)] [[PubMed](#)]
24. Donze, D.; Townes, T.M.; Bieker, J.J. Role of Erythroid Kruppel-like Factor in Human γ -to β -Globin Gene Switching (*). *J. Biol. Chem.* **1995**, *270*, 1955–1959. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, G.; Zhang, H.; Lin, A.; Wu, Z.; Li, T.; Zhang, X.; Chen, H.; Lu, D. Multi-Omics Analysis in β -Thalassemia Using an HBB Gene-Knockout Human Erythroid Progenitor Cell Model. *Int. J. Mol. Sci.* **2022**, *23*, 2807. [[CrossRef](#)] [[PubMed](#)]
26. Wang, H.; Hu, H.; Zhang, Q.; Yang, Y.; Li, Y.; Hu, Y.; Ruan, X.; Yang, Y.; Zhang, Z.; Shu, C. Dynamic transcriptomes of human myeloid leukemia cells. *Genomics* **2013**, *102*, 250–256. [[CrossRef](#)] [[PubMed](#)]
27. Zhu, X.; Wang, Y.; Pi, W.; Liu, H.; Wickrema, A.; Tuan, D. NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human γ -globin gene. *PLoS ONE* **2012**, e47175. [[CrossRef](#)]
28. Doerfler, P.A.; Feng, R.; Li, Y.; Palmer, L.E.; Porter, S.N.; Bell, H.W.; Crossley, M.; Pruett-Miller, S.M.; Cheng, Y.; Weiss, M.J. Activation of γ -globin gene expression by GATA1 and NF-Y in hereditary persistence of fetal hemoglobin. *Nat. Genet.* **2021**, *53*, 1177–1186. [[CrossRef](#)]
29. Gee, B.E.; Pearson, A.; Buchanan-Perry, I.; Simon, R.P.; Archer, D.R.; Meller, R. Whole Blood Transcriptome Analysis in Children with Sickle Cell Anemia. *Front. Genet.* **2021**, *12*, 737741. [[CrossRef](#)]
30. Wakil, S.M.; Alhissi, S.; Al Dossari, H.; Alqahtani, A.; Shibin, S.; Melaiki, B.T.; Finsterer, J.; Al-Hashem, A.; Bohlega, S.; Alazami, A.M. Truncating ARL6IP1 variant as the genetic cause of fatal complicated hereditary spastic paraplegia. *BMC Med Genet.* **2019**, *20*, 1–6. [[CrossRef](#)]
31. Nizon, M.; Küry, S.; Péréon, Y.; Besnard, T.; Quinquis, D.; Boisseau, P.; Marsaud, T.; Magot, A.; Mussini, J.M.; Mayrargue, E. ARL6IP1 mutation causes congenital insensitivity to pain, acromutilation and spastic paraplegia. *Clin. Genet.* **2018**, *93*, 169–172. [[CrossRef](#)] [[PubMed](#)]
32. Di Lisi, R.; Picard, A.; Ausoni, S.; Schiaffino, S. GATA elements control repression of cardiac troponin I promoter activity in skeletal muscle cells. *BMC Mol. Biol.* **2007**, *8*, 1–8.
33. Musarò, A.; McCullagh, K.J.; Naya, F.J.; Olson, E.N.; Rosenthal, N. IGF-1 induces skeletal myocyte hypertrophy through calcineurin in association with GATA-2 and NF-ATc1. *Nature* **1999**, *400*, 581–585. [[CrossRef](#)] [[PubMed](#)]
34. Paul, A.C.; Rosenthal, N. Different modes of hypertrophy in skeletal muscle fibers. *J. Cell Biol.* **2002**, *156*, 751–760. [[CrossRef](#)]
35. Lin, M.; Lin, J.; Hsu, C.; Juan, H.; Lou, P.; Huang, M. GATA3 interacts with and stabilizes HIF-1 α to enhance cancer cell invasiveness. *Oncogene* **2017**, *36*, 4243–4252. [[CrossRef](#)]
36. Remels, A.; Langen, R.; Gosker, H.R.; Russell, A.; Spaapen, F.; Voncken, J.; Schrauwen, P.; Schols, A.M. PPAR γ inhibits NF- κ B-dependent transcriptional activation in skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **2009**, *297*, E174–E183. [[CrossRef](#)]
37. Phua, W.W.T.; Wong, M.X.Y.; Liao, Z.; Tan, N.S. An aPPARent functional consequence in skeletal muscle physiology via peroxisome proliferator-activated receptors. *Int. J. Mol. Sci.* **2018**, *19*, 1425. [[CrossRef](#)]
38. De Rooij, J.D.; Beuling, E.; van den Heuvel-Eibrink, M.M.; Obulkasim, A.; Baruchel, A.; Trka, J.; Reinhardt, D.; Sonneveld, E.; Gibson, B.E.; Pieters, R. Recurrent deletions of IKZF1 in pediatric acute myeloid leukemia. *Haematologica* **2015**, *100*, 1151. [[CrossRef](#)]
39. Leng, D.; Miao, R.; Huang, X.; Wang, Y. In silico analysis identifies CRISP3 as a potential peripheral blood biomarker for multiple myeloma: From data modeling to validation with RT-PCR. *Oncol. Lett.* **2018**, *15*, 5167–5174. [[CrossRef](#)]
40. Pfisterer, P.; König, H.; Hess, J.; Lipowsky, G.; Haendler, B.; Schleuning, W.-D.; Wirth, T. CRISP-3, a protein with homology to plant defense proteins, is expressed in mouse B cells under the control of Oct2. *Mol. Cell. Biol.* **1996**, *16*, 6160–6168. [[CrossRef](#)]

41. Ng, A.P.; Coughlan, H.D.; Hediye-Zadeh, S.; Behrens, K.; Johanson, T.M.; Low, M.S.Y.; Bell, C.C.; Gilan, O.; Chan, Y.-C.; Kueh, A.J. An Erg-driven transcriptional program controls B cell lymphopoiesis. *Nat. Commun.* **2020**, *11*, 1–14. [[CrossRef](#)] [[PubMed](#)]
42. Knief, J.; Reddemann, K.; Gliemroth, J.; Brede, S.; Bartscht, T.; Thorns, C. ERG expression in multiple myeloma—A potential diagnostic pitfall. *Pathol. Res. Pract.* **2017**, *213*, 130–132. [[CrossRef](#)] [[PubMed](#)]
43. Tsuzuki, S.; Taguchi, O.; Seto, M. Promotion and maintenance of leukemia by ERG. *Blood J. Am. Soc. Hematol.* **2011**, *117*, 3858–3868. [[CrossRef](#)] [[PubMed](#)]
44. Koczera, P.; Martin, L.; Marx, G.; Schuerholz, T. The ribonuclease a superfamily in humans: Canonical RNases as the buttress of innate immunity. *Int. J. Mol. Sci.* **2016**, *17*, 1278. [[CrossRef](#)] [[PubMed](#)]
45. Parakati, R.; DiMario, J.X. Sp1-and Sp3-mediated transcriptional regulation of the fibroblast growth factor receptor 1 gene in chicken skeletal muscle cells. *J. Biol. Chem.* **2002**, *277*, 9278–9285. [[CrossRef](#)]
46. Irrcher, I.; Hood, D.A. Regulation of Egr-1, SRF, and Sp1 mRNA expression in contracting skeletal muscle cells. *J. Appl. Physiol.* **2004**, *97*, 2207–2213. [[CrossRef](#)]
47. Cifuentes-Diaz, C.; Delaporte, C.; Dautréaux, B.; Charron, D.; Fardeau, M. Class II MHC antigens in normal human skeletal muscle. *Muscle Nerve Off. J. Am. Assoc. Electrodiagn. Med.* **1992**, *15*, 295–302. [[CrossRef](#)]
48. Englund, P.; Lindroos, E.; Nennesmo, I.; Klareskog, L.; Lundberg, I.E. Skeletal muscle fibers express major histocompatibility complex class II antigens independently of inflammatory infiltrates in inflammatory myopathies. *Am. J. Pathol.* **2001**, *159*, 1263–1273. [[CrossRef](#)]
49. Van Den Elsen, P.J. Expression regulation of major histocompatibility complex class I and class II encoding genes. *Front. Immunol.* **2011**, *2*, 48. [[CrossRef](#)]
50. Castelli, E.C.; Veiga-Castelli, L.C.; Yaghi, L.; Moreau, P.; Donadi, E.A. Transcriptional and posttranscriptional regulations of the HLA-G gene. *J. Immunol. Res.* **2014**, *2014*, 734068. [[CrossRef](#)]
51. Carli, L.; Tani, C.; Vagnani, S.; Signorini, V.; Mosca, M. Leukopenia, lymphopenia, and neutropenia in systemic lupus erythematosus: Prevalence and clinical impact—A systematic literature review. In *Seminars in Arthritis and Rheumatism*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 45, pp. 190–194.
52. Bitencourt, N.; Solow, E.B.; Wright, T.; Bermas, B.L. Inflammatory myositis in systemic lupus erythematosus. *Lupus* **2020**, *29*, 776–781. [[CrossRef](#)] [[PubMed](#)]
53. Cheung, S.M.; Keenan, K.; Senn, N.; Hutcheon, G.; Chan, K.; Erwig, L.; Schrepf, A.; Dospinescu, P.; Gray, S.; Waiter, G. Metabolic and Structural Skeletal Muscle Health in Systemic Lupus Erythematosus-Related Fatigue: A Multimodal Magnetic Resonance Imaging Study. *Arthritis Care Res.* **2019**, *71*, 1640–1646. [[CrossRef](#)] [[PubMed](#)]
54. Vechetti, I.J., Jr.; Valentino, T.; Mobley, C.B.; McCarthy, J.J. The role of extracellular vesicles in skeletal muscle and systematic adaptation to exercise. *J. Physiol.* **2021**, *599*, 845–861. [[CrossRef](#)] [[PubMed](#)]
55. Rome, S.; Forterre, A.; Mizgier, M.L.; Bouzakri, K. Skeletal muscle-released extracellular vesicles: State of the art. *Front. Physiol.* **2019**, 929. [[CrossRef](#)]
56. Lu, Z.; Joseph, D.; Bugnard, E.; Zaal, K.J.; Ralston, E. Golgi complex reorganization during muscle differentiation: Visualization in living cells and mechanism. *Mol. Biol. Cell* **2001**, *12*, 795–808. [[CrossRef](#)]
57. Ralston, E.; Ploug, T.; Kalthovde, J.; Lomo, T. Golgi complex, endoplasmic reticulum exit sites, and microtubules in skeletal muscle fibers are organized by patterned activity. *J. Neurosci.* **2001**, *21*, 875–883. [[CrossRef](#)]
58. Donkervoort, S.; Krause, N.; Dergai, M.; Yun, P.; Koliwer, J.; Gorokhova, S.; Geist Hauserman, J.; Cummings, B.B.; Hu, Y.; Smith, R. BET1 variants establish impaired vesicular transport as a cause for muscular dystrophy with epilepsy. *EMBO Mol. Med.* **2021**, *13*, e13787. [[CrossRef](#)]
59. Buas, M.F.; Kadesch, T. Regulation of skeletal myogenesis by Notch. *Exp. Cell Res.* **2010**, *316*, 3028–3033. [[CrossRef](#)]
60. Raj, P.; Rai, E.; Song, R.; Khan, S.; Wakeland, B.E.; Viswanathan, K.; Arana, C.; Liang, C.; Zhang, B.; Dozmorov, I. Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *eLife* **2016**, *5*, e12089. [[CrossRef](#)]
61. Carey, B.S.; Poulton, K.V.; Poles, A. Factors affecting HLA expression: A review. *Int. J. Immunogenet.* **2019**, *46*, 307–320. [[CrossRef](#)]
62. Bronson, P.G.; Caillier, S.; Ramsay, P.P.; McCauley, J.L.; Zuvich, R.L.; De Jager, P.L.; Rioux, J.D.; Ivinson, A.J.; Compston, A.; Hafler, D.A. CIITA variation in the presence of HLA-DRB1* 1501 increases risk for multiple sclerosis. *Hum. Mol. Genet.* **2010**, *19*, 2331–2340. [[CrossRef](#)] [[PubMed](#)]
63. Gyllenberg, A.; Piehl, F.; Alfredsson, L.; Hillert, J.; Bomfim, I.L.; Padyukov, L.; Orho-Melander, M.; Lindholm, E.; Landin-Olsson, M.; Lernmark, Å. Variability in the CIITA gene interacts with HLA in multiple sclerosis. *Genes Immun.* **2014**, *15*, 162–167. [[CrossRef](#)] [[PubMed](#)]
64. Gyllenberg, A.; Asad, S.; Piehl, F.; Swanberg, M.; Padyukov, L.; Van Yserloo, B.; Rutledge, E.A.; McNeney, B.; Graham, J.; Orho-Melander, M. Age-dependent variation of genotypes in MHC II transactivator gene (CIITA) in controls and association to type 1 diabetes. *Genes Immun.* **2012**, *13*, 632–640. [[CrossRef](#)]
65. Nguyen, C.; Varney, M.D.; Harrison, L.C.; Morahan, G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. *Diabetes* **2013**, *62*, 2135–2140. [[CrossRef](#)]
66. Williams, R.; Muller, Y.; Hanson, R.; Knowler, W.; Mason, C.; Bian, L.; Ossowski, V.; Wiedrich, K.; Chen, Y.; Marcovina, S. HLA-DRB1 reduces the risk of type 2 diabetes mellitus by increased insulin secretion. *Diabetologia* **2011**, *54*, 1684–1692. [[CrossRef](#)] [[PubMed](#)]

67. Chinniah, R.; Sevak, V.; Pandi, S.; Ravi, P.M.; Vijayan, M.; Kannan, A.; Karuppiyah, B. HLA-DRB1 genes and the expression dynamics of HLA CIITA determine the susceptibility to T2DM. *Immunogenetics* **2021**, *73*, 291–305. [[CrossRef](#)]
68. DeFronzo, R.A.; Tripathy, D. Skeletal muscle insulin resistance is the primary defect in type 2 diabetes. *Diabetes Care* **2009**, *32* (Suppl. 2), S157–S163. [[CrossRef](#)]
69. Bouzakri, K.; Koistinen, H.A.; Zierath, J.R. Molecular mechanisms of skeletal muscle insulin resistance in type 2 diabetes. *Curr. Diabetes Rev.* **2005**, *1*, 167–174. [[CrossRef](#)]
70. Israni, N.; Goswami, R.; Kumar, A.; Rani, R. Interaction of vitamin D receptor with HLA DRB1* 0301 in type 1 diabetes patients from North India. *PLoS ONE* **2009**, *4*, e8023. [[CrossRef](#)]
71. Cramer, L.A.; Nelson, S.L.; Klemsz, M.J. Synergistic induction of the Tap-1 gene by IFN- γ and lipopolysaccharide in macrophages is regulated by STAT1. *J. Immunol.* **2000**, *165*, 3190–3197. [[CrossRef](#)]
72. White, L.C.; Wright, K.L.; Felix, N.J.; Ruffner, H.; Reis, L.F.; Pine, R.; Ting, J.P.-Y. Regulation of LMP2 and TAP1 genes by IRF-1 explains the paucity of CD8+ T cells in IRF-1 $^{-/-}$ mice. *Immunity* **1996**, *5*, 365–376. [[CrossRef](#)]
73. Mogensen, T.H. IRF and STAT transcription factors—from basic biology to roles in infection, protective immunity, and primary immunodeficiencies. *Front. Immunol.* **2019**, *9*, 3047. [[CrossRef](#)] [[PubMed](#)]
74. Fink, K.; Grandvaux, N. STAT2 and IRF9: Beyond ISGF3. *Jak-Stat* **2013**, *2*, e27521. [[CrossRef](#)] [[PubMed](#)]
75. Rengachari, S.; Groiss, S.; Devos, J.M.; Caron, E.; Grandvaux, N.; Panne, D. Structural basis of STAT2 recognition by IRF9 reveals molecular insights into ISGF3 function. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E601–E609. [[CrossRef](#)]
76. McNab, F.; Mayer-Barber, K.; Sher, A.; Wack, A.; O’garra, A. Type I interferons in infectious disease. *Nat. Rev. Immunol.* **2015**, *15*, 87–103. [[CrossRef](#)]
77. Hemann, E.A.; Gale, M., Jr.; Savan, R. Interferon lambda genetics and biology in regulation of viral control. *Front. Immunol.* **2017**, *8*, 1707. [[CrossRef](#)]
78. Rivera, A. Interferon lambda’s new role as regulator of neutrophil function. *J. Interferon Cytokine Res.* **2019**, *39*, 609–617. [[CrossRef](#)]
79. Paul, A.; Tang, T.H.; Ng, S.K. Interferon regulatory factor 9 structure and regulation. *Front. Immunol.* **2018**, *9*, 1831. [[CrossRef](#)]
80. Cheon, H.; Holvey-Bates, E.G.; Schoggins, J.W.; Forster, S.; Hertzog, P.; Imanaka, N.; Rice, C.M.; Jackson, M.W.; Junk, D.J.; Stark, G.R. IFN β -dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. *EMBO J.* **2013**, *32*, 2751–2763. [[CrossRef](#)]
81. Blaszczyk, K.; Olejnik, A.; Nowicka, H.; Ozgyin, L.; Chen, Y.-L.; Chmielewski, S.; Kostyrko, K.; Wesoly, J.; Balint, B.L.; Lee, C.-K. STAT2/IRF9 directs a prolonged ISGF3-like transcriptional response and antiviral activity in the absence of STAT1. *Biochem. J.* **2015**, *466*, 511. [[CrossRef](#)]
82. Thibault, D.L.; Chu, A.D.; Graham, K.L.; Balboni, I.; Lee, L.Y.; Kohlmoos, C.; Landrigan, A.; Higgins, J.P.; Tibshirani, R.; Utz, P.J. IRF9 and STAT1 are required for IgG autoantibody production and B cell expression of TLR7 in mice. *J. Clin. Investig.* **2008**, *118*, 1417–1426. [[CrossRef](#)] [[PubMed](#)]
83. Goropevšek, A.; Gorenjak, M.; Gradišnik, S.; Dai, K.; Holc, I.; Hojs, R.; Krajnc, I.; Pahor, A.; Avčín, T. Increased levels of STAT1 protein in blood CD4 T cells from systemic lupus erythematosus patients are associated with perturbed homeostasis of activated CD45RA-FOXP3hi regulatory subset and follow-up disease severity. *J. Interferon Cytokine Res.* **2017**, *37*, 254–268. [[CrossRef](#)] [[PubMed](#)]
84. Wang, J.; Dai, M.; Cui, Y.; Hou, G.; Deng, J.; Gao, X.; Liao, Z.; Liu, Y.; Meng, Y.; Wu, L. Association of Abnormal Elevations in IFIT 3 With Overactive Cyclic GMP-AMP Synthase/Stimulator of Interferon Genes Signaling in Human Systemic Lupus Erythematosus Monocytes. *Arthritis Rheumatol.* **2018**, *70*, 2036–2045. [[CrossRef](#)] [[PubMed](#)]
85. Liu, W.; Li, M.; Wang, Z.; Wang, J. Ifn- γ mediates the development of systemic lupus erythematosus. *BioMed Res. Int.* **2020**, *2020*, 7176515. [[CrossRef](#)] [[PubMed](#)]
86. Wang, J.; Gamazon, E.R.; Pierce, B.L.; Stranger, B.E.; Im, H.K.; Gibbons, R.D.; Cox, N.J.; Nicolae, D.L.; Chen, L.S. Imputing gene expression in uncollected tissues within and beyond GTEx. *Am. J. Hum. Genet.* **2016**, *98*, 697–708. [[CrossRef](#)]
87. Lopes-Ramos, C.M.; Paulson, J.N.; Chen, C.-Y.; Kuijjer, M.L.; Fagny, M.; Platig, J.; Sonawane, A.R.; DeMeo, D.L.; Quackenbush, J.; Glass, K. Regulatory network changes between cell lines and their tissues of origin. *BMC Genom.* **2017**, *18*, 723. [[CrossRef](#)]
88. Devenish, L.P.; Mhlanga, M.M.; Negishi, Y. Immune Regulation in Time and Space: The Role of Local-and Long-Range Genomic Interactions in Regulating Immune Responses. *Front. Immunol.* **2021**, *12*, 1715. [[CrossRef](#)]
89. Smale, S.T. Transcriptional regulation in the immune system: A status report. *Trends Immunol.* **2014**, *35*, 190–194. [[CrossRef](#)]
90. helu2008/transTFModel. Available online: <https://github.com/helu2008/transTFModel> (accessed on 4 August 2021).