



Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrmTransfer of regulatory knowledge from human to mouse for functional genomics analysis[☆]Christian H. Holland^{a,b}, Bence Szalai^c, Julio Saez-Rodriguez^{a,b,*}^a Institute of Computational Biomedicine, Heidelberg University, Faculty of Medicine, Bioquant - Im Neuenheimer Feld 267, 69120 Heidelberg, Germany^b Joint Research Centre for Computational Biomedicine (JRC-COMBINE), RWTH Aachen University, Faculty of Medicine, Pauwelsstrasse 19, 52074 Aachen, Germany^c Department of Physiology, Semmelweis University, Faculty of Medicine, Tűzoltó u. 37-47, 1085 Budapest, Hungary

ARTICLE INFO

Keywords:

Functional genomics

Signaling footprints

Pathway activity

Transcription factor activity

ABSTRACT

Transcriptome profiling followed by differential gene expression analysis often leads to lists of genes that are hard to analyze and interpret. Functional genomics tools are powerful approaches for downstream analysis, as they summarize the large and noisy gene expression space into a smaller number of biological meaningful features. In particular, methods that estimate the activity of processes by mapping transcripts level to process members are popular. However, footprints of either a pathway or transcription factor (TF) on gene expression show superior performance over mapping-based gene sets. These footprints are largely developed for humans and their usability in the broadly-used model organism *Mus musculus* is uncertain. Evolutionary conservation of the gene regulatory system suggests that footprints of human pathways and TFs can functionally characterize mice data. In this paper we analyze this hypothesis. We perform a comprehensive benchmark study exploiting two state-of-the-art footprint methods, DoRotheA and an extended version of PROGENy. These methods infer TF and pathway activity, respectively. Our results show that both can recover mouse perturbations, confirming our hypothesis that footprints are conserved between mice and humans. Subsequently, we illustrate the usability of PROGENy and DoRotheA by recovering pathway/TF-disease associations from newly generated disease sets. Additionally, we provide pathway and TF activity scores for a large collection of human and mouse perturbation and disease experiments (2374). We believe that this resource, available for interactive exploration and download (https://saezlab.shinyapps.io/footprint_scores/), can have broad applications including the study of diseases and therapeutics.

1. Introduction

The typical framework of functional genomics studies comprises the analysis of expression changes of groups of genes. These groups are referred to as gene sets and typically consist of genes sharing common functions (e.g. Gene Ontology analysis) or genes encoding for pathway members [1]. The latter are used for classical pathway analysis studies, which assume that the transcript level is a proxy for protein abundance and thus the pathway activity. The framework of estimating transcription factor (TF) activity based on its gene expression follows the same principle (Fig. 1A). However, mapping the transcript level to proteins neglects the effects of post-transcriptional and post-translational modifications, even though they are essential for the function of

many proteins [2].

To overcome this limitation, alternative approaches have been developed which are based on newly derived gene sets containing gene signatures obtained from genetic or chemical perturbations of pathways or TFs. These signatures are the footprint on gene expression of the corresponding pathway or TF (Fig. 1A). Recent studies indicate that footprints outperform mapping gene sets [3,4]. Since most of these footprints are generated for the application in humans, their usability in model organisms is uncertain. This question is of importance since the study of human diseases is limited by the availability of patient data and ethical concerns, and are thus often complemented with experimental work in model organisms, in particular mice (*Mus musculus*) [5].

Perturbation of gene expression in humans can be estimated from

[☆] This article is part of a Special Issue entitled: Transcriptional Profiles and Regulatory Gene Networks edited by Dr. Dr. Federico Manuel Giorgi and Dr. Shaun Mahony.

* Corresponding author at: Institute of Computational Biomedicine, Heidelberg University, Faculty of Medicine, Bioquant - Im Neuenheimer Feld 267, 69120 Heidelberg, Germany.

E-mail addresses: christian.holland@bioquant.uni-heidelberg.de (C.H. Holland), szalai.bence@med.semmelweis-univ.hu (B. Szalai), julio.saez@bioquant.uni-heidelberg.de (J. Saez-Rodriguez).

<https://doi.org/10.1016/j.bbagrm.2019.194431>

Received 31 May 2019; Received in revised form 31 August 2019; Accepted 1 September 2019

1874-9399/ © 2019 Elsevier B.V. All rights reserved.

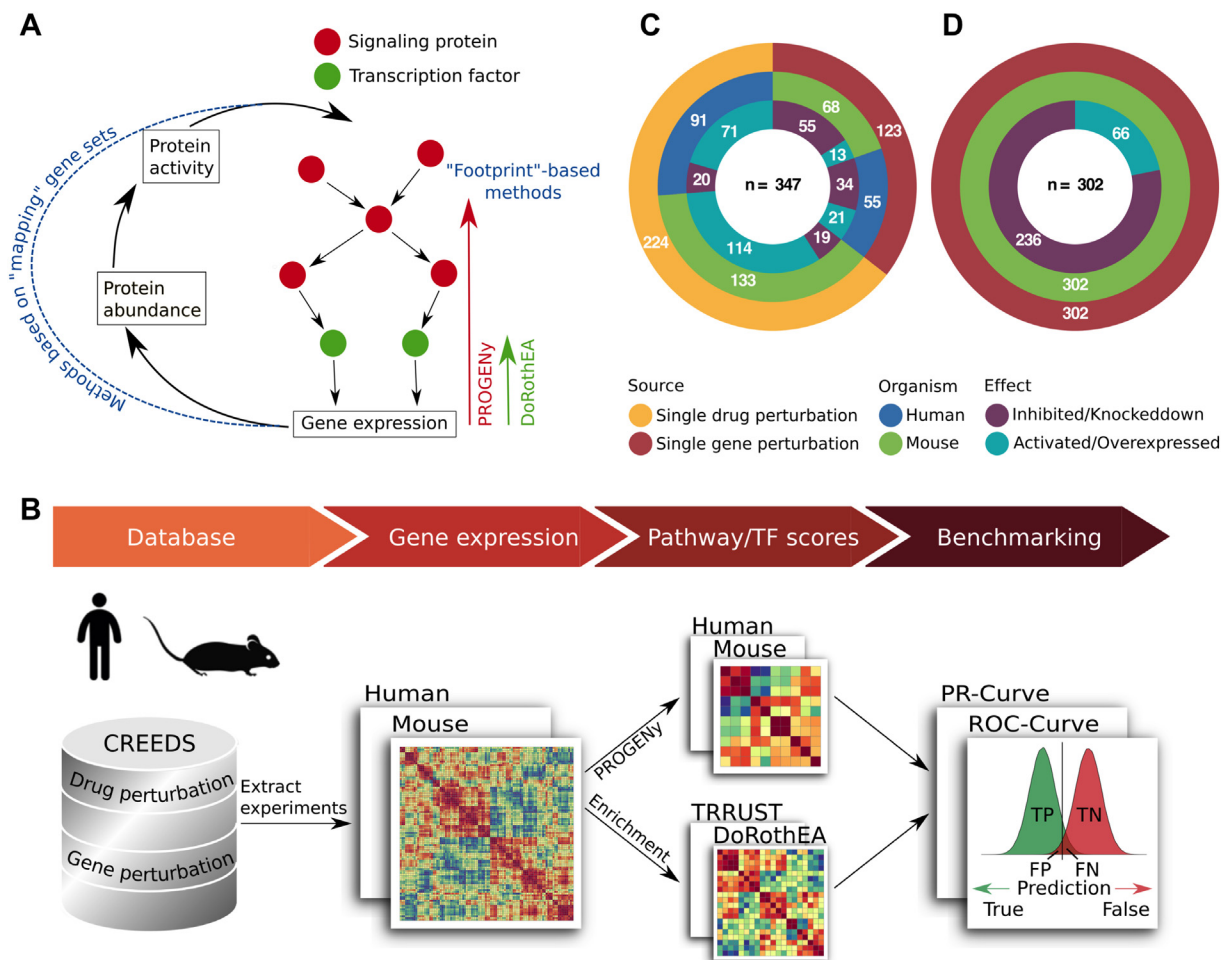


Fig. 1. Benchmark overview. (A) Visualization of the classical 'mapping' framework, where transcript level is mapped to protein level and thus to protein activity in contrast to the footprint based methods PROGENy and DoRothEA. (B) Benchmark pipeline starting with the extraction of mouse and human single gene and single drug perturbation experiments from the CREEDS database. Pathway and TF activities are computed for each experiment separately based on their differential expression signature. For the PROGENy benchmark we compared human-PROGENy vs mouse-PROGENy. For DoRothEA benchmark we compared mouse-DoRothEA vs dedicated mouse regulons from TRRUST. We evaluate the performance of both approaches using ROC and PR-metrics. (C) Overview of benchmark datasets for PROGENy (C) and DoRothEA (D), including the perturbation type, organism, and perturbation effect. Numbers indicate the amount of experiments in each group.

mouse transcriptomic data [6,7]. Furthermore, previous studies suggest that pathway and TF footprints are evolutionarily conserved between mice and humans: pathway footprints derived from mouse B cells can provide valuable insights into human cancer [8], and inferred prostate-specific gene regulatory networks of mice and humans overlap by over 70% [9]. This suggests that human functional genomics tools, which consider footprints as gene sets, could be applied on mice data. However, as of today there is no comprehensive study to prove this.

To validate whether pathway and TF footprints are evolutionarily conserved between mice and humans we performed a comprehensive benchmark study. We exploited two state of the art functional genomics approaches covering both aspects of gene regulatory networks: signaling pathways and transcriptional regulation. The first approach is PROGENy, a tool that estimated the activity of, originally, 11 signaling pathways from gene expression data [3]. It is based on consensus transcriptomic perturbation signatures, we refer to as footprints, of signaling pathways on gene expression. In this work we extended PROGENy with novel footprints of the signaling pathways Androgen, Estrogen, and WNT. The second approach is DoRothEA, a resource matching TFs with their transcriptional targets [10], which allows us to estimate TF activity from gene expression data in humans by enriched regulon analysis [11]. We consider the targets of a TF also as footprints of a TF on gene expression. We validated that both PROGENy and DoRothEA can recover mice perturbations, supporting our hypothesis

about the conserved nature of pathway and TF footprints. To demonstrate the usability of PROGENy and DoRothEA we estimated pathway and TF activities for a large collection of mice and human diseases as well as chemical and genetic perturbation experiments. Based on the activities of the disease experiments we were able to recover known pathway/TF disease associations. For this, we constructed 738 novel disease sets matching 186 diseases with 467 disease experiments.

2. Methods

2.1. Benchmark dataset collection

We collected transcriptomic perturbation experiments in human and mouse profiled by single-channel microarrays from the database CREEDS [12], which contains among others resources single drug and single gene perturbation experiments from Gene Expression Omnibus (GEO). We extended this collection with manually curated perturbation experiments from ArrayExpress using the data collection and curation pipeline described previously [3]. Arrays with no raw data available or with no corresponding annotation package were discarded. We translated GEO accession ids to ArrayExpress accession ids and downloaded CEL files for all experiments using the function ArrayExpress from the BioConductor package ArrayExpress (version 1.40.0) [13].

2.2. New PROGENy pathways

PROGENy is based on footprints which are consensus gene signatures delivered from pathway-related perturbation experiments. We added 3 new pathways footprints (Androgen, Estrogen and WNT) to the existing 11 pathways of PROGENy in this study. To collect the corresponding perturbation experiments, we queried ArrayExpress [13] with keywords {'androgen', 'DHT', 'testosterone'}, {'estrogen', 'SERM', 'tamoxifen'} and {'APC', 'axin', 'catenin', 'Frizzled', 'GSK3', 'WNT'} for Androgen, Estrogen and WNT pathways, respectively. For further curation and PROGENy model fitting we used the pipeline described previously [3]. Accession number of the finally used experiments is given in Supplementary Table S1.

2.3. Microarray processing

The processing steps from raw data to annotated probe levels, comprising quality control, background correction, normalization and annotation is described in the original PROGENy paper [3].

Experiments with less than two control replicates remaining after the processing step were discarded. We used the BioConductor package limma (version 3.36.2) [14] to perform differential expression analysis calculating the contrast between perturbed and control replicates. Instead of log-fold changes we considered the moderated t-value as gene-level statistic.

2.4. Construction of mouse-PROGENy and calculation of pathway activities

The original PROGENy model is a matrix with footprint genes in rows and pathways in columns. The entries denote a measure accounting for how genes respond to pathway perturbation (up- or downregulation). Each pathway is limited to the top 100 most responsive genes.

To construct mouse-PROGENy we mapped the HGNC symbols of the original PROGENy matrix to their ortholog MGI-symbol using the BioConductor package biomaRt (version 2.36.1, ensembl release 96 April 2019) [15]. The mapping can lead to duplicated genes. Either a single HGNC symbol is mapped to several MGI symbols or several HGNC symbols are mapped to a single MGI symbol. In the first case, the weight of the HGNC symbol is divided by the number of mapping MGI genes. In the second case, the weight of the new MGI symbol is the arithmetic mean value of all mapping HGNC symbols. Following this strategy we introduce a bias to human genes which have multiple orthologs in mice (e.g. the gene SERPINA3 that belongs to the footprint of JAK-STAT maps to 10 mouse genes: Serpina3a/b/c/f/g/i/j/k/m/n. Hence the JAK-STAT pathway is biased to SERPINA3). In order to compensate for this bias we extend the pathway footprints with a number of additional genes equal to the number of ortholog genes per human gene - 1. Finally, a mouse specific PROGENy matrix is retrieved (mouse-PROGENy) so that we can estimate pathway activity scores from mice gene expression data for the original 11 (EGFR, Hypoxia, JAK-STAT, MAPK, NFkB, p53, PI3K, TGFb, TNFa, Trail, VEGF) and the 3 newly added (Androgen, Estrogen, WNT) pathways. The pathway activity scores are calculated for each contrast using moderated t-values as gene-level statistic. Activity scores are pathway-wise z-score normalized.

2.5. Construction of mouse-DoRothEA and calculation of TF activities

DoRothEA is a resource linking human TFs to their direct transcriptional targets. Each TF is accompanied with a summary confidence level from A (high confidence) to E (low confidence) based on the amount of supporting TF's regulatory evidence. We inferred mouse-DoRothEA by mapping HGNC-symbols to their ortholog MGI-symbol using the BioConductor package BiomaRt (version 2.36.1, ensembl release 96 April 2019) [15]. The mapping can lead to TFs with multiple

confidence levels. To be more conservative, the lowest confidence level is chosen as TF-confidence level. The BioConductor package viper (version 1.14.0) [11] considers the regulons as gene sets and, thus, estimates TF activities from gene expression data using enriched regulon analysis. TF activity scores are computed for each contrast using moderated t-values as gene-level statistic. Only regulons comprised of at least 4 targets are tested. We consider the normalized enrichment score (NES) provided by viper as a measure for TF activity.

2.6. Quality control of single gene perturbation experiments

Single gene perturbation experiments provide the possibility for an intuitive quality control of the effect of the perturbation. If the gene-level statistic (t-value) sign of the perturbation target is not in agreement with the underlying perturbation ((+) for overexpression, (-) for knockdown/knockout) the perturbation experiment is considered unsuccessful and is thus discarded.

2.7. Computing ROC and PR curves

To transform the benchmark into a binary setup, all activity scores of experiments with negative perturbation effect (inhibition/knock-down) are multiplied by -1. This guarantees, that TFs/pathways belong to a binary class either deregulated or not regulated.

We computed the ROC-curves and associated statistics using the R package pROC (version 1.12.1) [16]. For PR-curves we used the R package PRROC (version 1.3.1) [17]. For the construction of ROC and PR curves we calculated for each perturbation experiment pathway (or TF) activities using PROGENy (or DoRothEA). As each perturbation experiment targets either a single pathway (or TF) only the activity score of the perturbed pathway (or TF) is associated with the positive class (e.g. EGFR pathway activity score in an experiment where EGFR was perturbed). Accordingly the activity scores of all non-perturbed pathways (or TFs) belong to the negative class (e.g. EGFR pathway activity score in an experiment where JAK-STAT pathway was perturbed). Using these positive and negative classes Sensitivity/(1-Specificity) or Precision/Recall values were calculated at different thresholds of activity, producing the ROC/PRC curves.

2.8. Downsampling true negatives

ROC curves are recommended when the numbers of true positives and true negatives are balanced [18]. In order to balance our benchmark dataset we downsampled the number of true negatives to equal the number of true positives 3000 times and computed AUROC for each run.

2.9. Inference of disease sets using disease ontology network

To create disease sets we determined all related parent diseases of the diseases studied in CREEDs by using the function ancestors from the BioConductor package rols (version 2.9.1) (<http://lgatto.github.com/rols/>) which provides an R interface to the Ontology Lookup service [19]. Each possible parent disease serves as a distinct disease set. CREEDs disease experiments which matches a child disease of a given disease set is considered as a set member.

2.10. Disease set enrichment analysis

To explore TF/Pathway-disease associations we downloaded all human and mouse disease signatures from the CREEDs database [12]. We followed the processing steps described before resulting in an expression vector of moderated t-values for each disease experiment. We computed pathway and TF activity scores for each vector. To apply the Gene Set Enrichment Analysis framework [1] with our disease sets we used the BioConductor package fgsea (version 1.6.0) [20]. Separately

for each pathway/TF (e.g. only for EGFR), disease experiments (with associated diseases) are ranked based on the pathway/TF activity score (e.g. the activity score of EGFR). Subsequently, it is tested whether experiments belonging to the same disease set are enriched either at the top or at the bottom of the list. Disease sets with less than 5 and more than 45 members were discarded. *P*-values were adjusted for multiple comparisons using false discovery rate (FDR) [21].

3. Results

3.1. Benchmark pipeline

We established a benchmark pipeline to discover whether both PROGENy and DoRothEA human footprint methods could be applied to functionally characterize mice data (Fig. 1B). Pathway/TF perturbation gene expression studies provide the opportunity to benchmark both tools: we can compare the predicted pathway and TF activities with the ‘ground truth’, denoted as the original perturbed target. The database CREEDS (CRowd Extracted Expression of Differential Signatures) provides manually curated single drug and single gene perturbation experiments in human and mice [12]. Additionally, we manually curated single drug perturbation experiments (see Methods). We included both perturbation directions - either activation/overexpression or inhibition/knockdown.

For the PROGENy validation we exploited both single drug and single gene perturbation studies. Experiments are considered to be relevant for our study if the perturbation target is a member or a gene encoding for a member of a PROGENy signaling pathway (drug/gene-PROGENy-pathway associations are provided in Supplementary Table S2). We identified 347 experiments (123 single gene and 224 single drug perturbation; Fig. 1C). These experiments cover 11 and 13 out of 14 possible pathways for human and mouse, respectively. These 14 pathways include Androgen, Estrogen, and WNT besides the 11 in the original PROGENy publication (see Methods) [3]. For DoRothEA we extracted only those single gene perturbation experiments where the target gene encodes for a TF which is defined by the human TF census from TFClass [22]. In total we collected 302 single gene perturbation experiments covering 144 mouse TFs (Fig. 1D).

To evaluate if PROGENy is applicable on mice data in the fairest way we would need to compare mouse-PROGENy vs a PROGENy version which was originally developed for application in mice. Since, to our knowledge, this resource does not currently exist we compared the performance of the newly derived mouse-PROGENy versus the original human-PROGENy tool. Note that this procedure introduces a bias towards the benchmark data as we benchmark mouse-PROGENy and human-PROGENy on independent data sets: one with human and the other with mouse pathway perturbation experiments. Regarding DoRothEA, there are resources that provide TF-target interactions for mice. Hence, we compared the newly derived mouse-DoRothEA versus dedicated mouse regulons from the TRRUST database [23]. Both mouse-PROGENy and mouse-DoRothEA were constructed by mapping human genes to their orthologs in mice (see Methods). To assess the model's prediction power we utilized the Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves (see Methods).

3.2. Benchmarking PROGENy

To compare mouse-PROGENy and human-PROGENy unbiasedly, we included only pathways perturbed in both benchmark datasets. Moreover, we evaluated PROGENy's global performance across all pathways. Both models clearly performed better than random (AUROC of 0.71 with 95% confidence interval of 0.662–0.757 and AUROC of 0.656 with 95% confidence interval of 0.610–0.703 for human and mouse-PROGENy respectively; Fig. 2A; ROC-curves for each pathway in Supplementary Fig. S1; summarized performance metrics are available in Supplementary Table S3). AUROC was not significantly different

between mouse and human (DeLong-test, $p = 0.113$). As our benchmark dataset is imbalanced (10% belong to the positive class) we also computed AUROC's upon multiple downsampling true negatives to equal the number of true positives with a resulting median AUROC equal to the AUROC of the unbalanced dataset (see Methods; Supplementary Fig. S2A and B). With precision-recall analysis, we obtained consistent results: human-PROGENy performed comparably to mouse-PROGENy (AUPRC of 0.254 and 0.246, respectively; Fig. 2B; PR-curves for each pathway are provided in Supplementary Fig. S3; summarized performance metrics are available in Supplementary Table S3). In addition, both performed better than a random model which would result in an AUPRC of 0.1. In summary, mouse-PROGENy performed comparably to human-PROGENy and better than a random model, regardless of the metric used. Thus, we conclude that PROGENy can recover pathway perturbations in mice.

3.3. Benchmarking DoRothEA

To evaluate if DoRothEA's regulons can functionally characterize mice data, we next compared the performance of mouse-DoRothEA to the performance of dedicated mouse regulons from the TRRUST database [23]. Human-DoRothEA was reconstructed by integrating different resources spanning from literature-curated databases to predictions of TF-target interactions [10]. Thereby, each TF is accompanied with a summary confidence level from A (high confidence) to E (low confidence) based on the amount of supporting TF's regulatory evidence. Our novel mouse-DoRothEA regulons comprise in total 1170 TFs, targeting 17,512 unique targets with 402,937 interactions distributed across all confidence levels (see Methods; Supplementary Fig. S4A). In contrast, TRRUST covers 828 TFs, which overlap 553 TFs from mouse-DoRothEA (Supplementary Fig. S4B). Comparing similarity of overlapping regulons between TRRUST and mouse-DoRothEA revealed, for most regulons, substantial differences (Supplementary Fig. S4C). To benchmark the performance of mouse-DoRothEA and TRRUST unbiasedly, we consider only the 553 TFs which are available in both resources. We cover 34–76 TFs of those 553 TFs (dependent on mouse-DoRothEA confidence level) with our benchmark data (Supplementary Fig. S4D). Moreover, we evaluated DoRothEA's global performance across all TFs since there were not enough public data sets available to evaluate the performance at the TF level. In ROC space mouse-DoRothEA outperformed TRRUST at any confidence level combination (Fig. 3A, summarized performance metrics are available in Supplementary Table S4). However, in PR space we found that that TRRUST has an advantage throughout all confidence level combinations except AB where DoRothEA is slightly better (Fig. 3B). All model subtypes performed better than a corresponding random model. In both metrics, we saw a peak at combined confidence level of A and B. Therefore, we decided to consider only TFs accompanied with the confidence levels A and B in the following analysis.

While both regulons resources performed better than random, mouse-DoRothEA (AUROC: 0.711, 95% confidence interval: 0.649–0.772) performed better than TRRUST (AUROC: 0.671, 95% confidence interval: 0.604–0.738; Fig. 3C), but without a significant difference (DeLong-test, $p = 0.249$). As the DoRothEA benchmark dataset is even more imbalanced (2.63% belong to the positive class) we downsampled again true negatives to equal the number of true positives, showing that the median of downsampled AUROC is equal to the AUROC of the imbalanced dataset (see Methods; Supplementary Fig. S5A and B). In PR-space mouse-DoRothEA performed just as well as TRRUST (AUPRC of 0.108 for both; Fig. 3D). Also both performed better than a random model with a corresponding AUPRC of 0.026. Considering the aforementioned results, we conclude that mouse-DoRothEA performs comparably to TRRUST and can thus recover transcriptional regulation in mice.

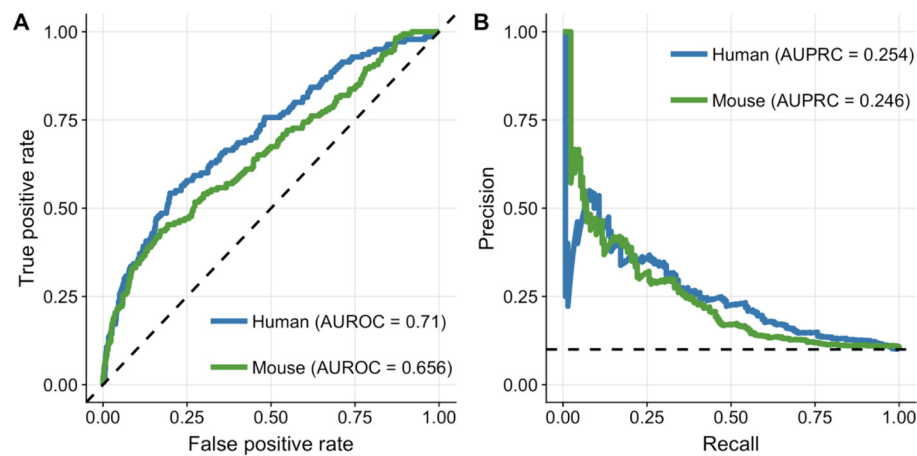


Fig. 2. Benchmark of mouse-PROGENy. ROC-curve (A) and PR-curve (B) analysis comparing human-PROGENy vs. mouse-PROGENy. Dashed lines indicate the performance of a random model.

3.4. Pathway/TF-disease associations

Once shown that PROGENy and DoRotheA can also be applied to mice data, we investigated whether we can recover known associations between pathways/TFs and human diseases based on transcriptomic disease signatures of both mice and humans. We downloaded 469 disease signatures from the CREEDs database [12] and computed pathway and TF activity scores for each experiment. To find associations we developed individual disease sets based on a disease ontology network from EBI Ontology Lookup Service [19]. Each node in this network represents a distinct disease set. If a node itself or a descendant of a node matched a CREEDs signature disease, the corresponding CREEDs

experiment is considered as a member of the disease set (Fig. 4A). In total we tested 732 distinct disease sets (Supplementary Material S1). Using these disease sets we found 434 significant (Gene set Enrichment Analysis [1]; $FDR \leq 0.1$ & $|NES| \geq 1$; see Methods) pathway-disease associations and 3586 significant ($FDR \leq 0.1$ & $|NES| \geq 1$) TF-disease associations covering 156 and 281 disease sets, respectively (Supplementary material 2).

The results were, in general, dominated by upregulated activity of two TFs, ETS2 (100 associations) and E2F1 (88 associations; Fig. 4B). Both are well-known oncogenes driving tumorigenesis [24,25]. Accordingly, most of their associations we found were related to different forms of cancer. Similarly, we found the activity of the tumor

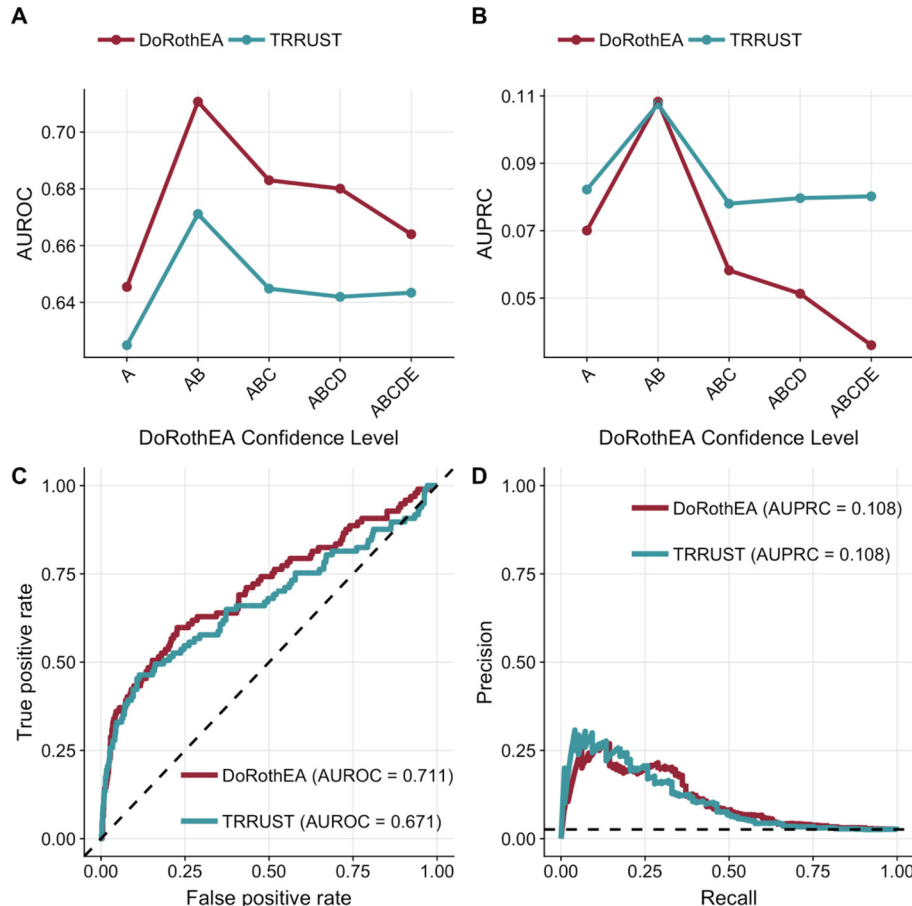


Fig. 3. Benchmark of mouse-DoRotheA. DoRotheA result of ROC-curve (A) and PR-curve (B) analysis summarized in AUROC and AUPRC, respectively, for different confidence level cutoffs. ROC-curve (C) and PR-curve (D) analysis comparing mouse-DoRotheA filtered for TFs with confidence level A or B vs. mouse-TRRUST. Dashed lines indicate the performance of a random model.

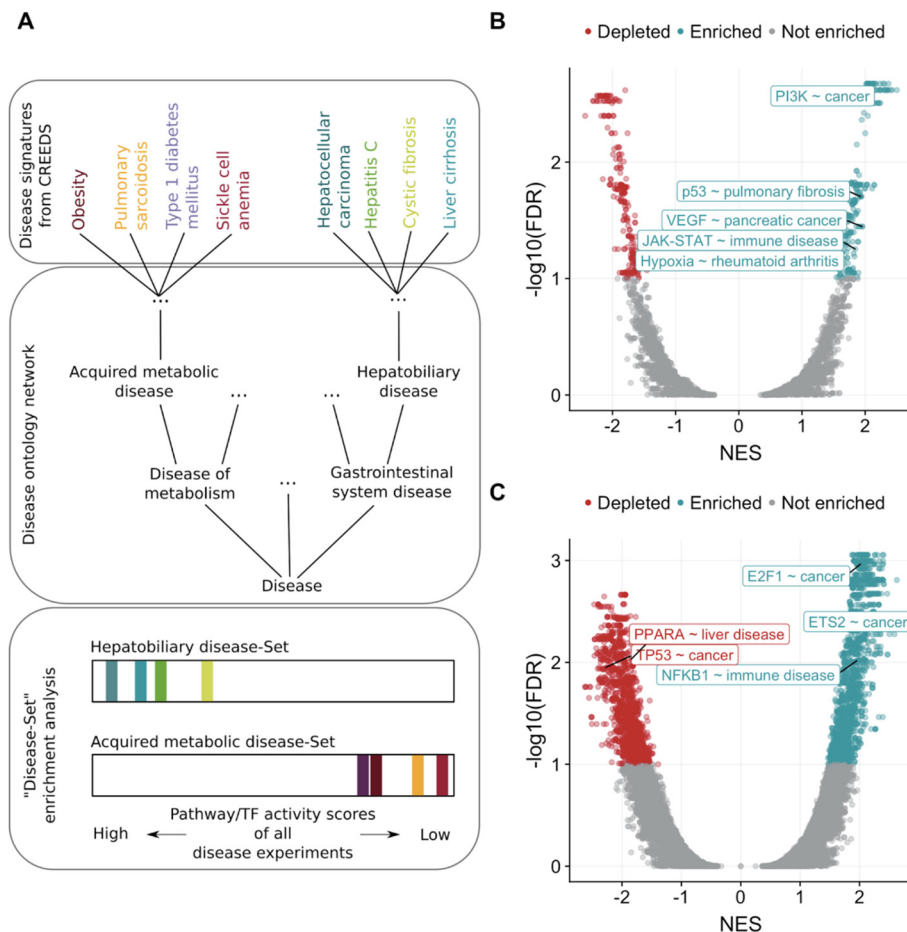


Fig. 4. Pathway/TF-disease associations. (A) Framework of gene set enrichment adapted for 'disease set' enrichment. The disease sets are created based on a disease ontology network. Each node in the network represents a disease set. CREDS diseases which are the node itself or descendants of a node are considered as corresponding disease set members. To perform the enrichment PROGENy and DoRothea activity levels are ranked separately and checked whether a disease set is enriched either at the top or at the bottom of the ranked list. (B, C) Volcano plots showing separately for pathways and TFs the outcome of disease set enrichment. Selected associations are labelled.

suppressor TP53 to be downregulated in cancer (7 associations). Pathway specific associations were dominated by the PI3K pathway (86 associations; Fig. 4C). Additionally, almost half of them were associations with different forms of cancer. Our approach revealed for the majority of all cancer associations an elevated activity level of PI3K. This finding is in agreement with the literature that describes PI3K as having control over important hallmarks of cancer, i.e. cell cycle, survival, and metabolism [26]. Also we found a strong upregulation of VEGF pathway in pancreatic cancer. Overexpressed VEGF (Vascular endothelial growth factor) is involved in angiogenesis and is considered as a diagnostic marker for pancreatic cancer [27]. These examples emphasize the importance of signaling pathways and transcriptional regulation in the context of cancer diseases.

However, next to cancer related diseases we also recovered strong associations with other disease types, e.g. upregulated Hypoxia pathway activity in rheumatoid arthritis [28]. Also, NFKB1 and JAK-STAT showed elevated activity in immune and, therefore, leukocyte related diseases, such as inflammation of the lung, bowel, mucous membrane, or skin [29,30].

In the context of chronic liver disease we recovered the role of PPARA. Its expression is reduced in hepatic stellate cells during liver fibrosis [31]. This finding is in agreement with our study as we found down regulated PPARA activity associated with the set 'liver disease'. Moreover, reduced PPARA activity was also significantly depleted within the disease sets 'hepatocellular carcinoma' and 'liver carcinoma'.

Altogether, we showed that PROGENy and DoRothea are capable to recover known signaling pathway/TF disease association based on mice and human data.

4. Discussion

The evolutionarily conserved gene regulatory system between mouse and human suggests that the footprints of a pathway or TF on gene expression are evolutionarily conserved as well. This hypothesis has a direct impact on footprint methods developed for human application, such as PROGENy and DoRothea. Both rely on gene sets comprising footprints and given that our assumption is true, they can be applied to mice data, which is an important resource for the study of human diseases. We addressed this question by establishing a benchmark pipeline to validate if DoRothea and an extended version of PROGENy (we added footprints for Androgen, Estrogen and WNT) can be applied to functionally characterize mice data (Fig. 1B).

We found that mouse-PROGENy is globally effective in inferring pathway activity on mouse data. However, the pathway-wise benchmark showed that the prediction power varies across pathways (Supplementary Figs. S1 and S3). Especially for JAK-STAT, we saw a highly significant difference between mice and humans in ROC space for the benefit of humans (DeLong-test, $p = 5e-5$). Interestingly, we observed the inverse case for the pathway NFKB. Here, mouse-PROGENy tends to outperform human-PROGENy (DeLong-test, $p = 0.057$), while NFKB still performed well in human [3]. This difference emphasizes the importance of the quality of the benchmark data. The benchmark data in [3] was curated very carefully by reviewing each perturbation experiment separately. Our analysis is based on a broad collection of curated experiments via crowdsourcing. By their own nature, crowdsourcing projects cannot be fully controlled, and miss annotations can occur, which could contribute to the low performance we found for some pathways. Other pathways, with a low number of positive cases such as VEGF, must be interpreted with caution as reliability of ROC/PR tests decreases with decreasing positive

cases.

Regarding mouse-DoRothEA, we found its performance comparable to dedicated mouse regulons from TRRUST. However, we recommend the use of mouse-DoRothEA instead of TRRUST as it provides a better coverage at similar performance. Regulons with confidence levels A and B have been shown to perform the best for both resources. Including confidence level C almost doubled the TF coverage from 34 to 59 TFs (Supplementary Fig. S4B) but caused a performance drop. By including TFs labelled with confidence level C, we introduce regulons in our benchmark data that have not been thoroughly studied [10], hence the drop in performance is expected.

Our above stated findings about the performance of PROGENy and DoRothEA support our initial hypothesis that footprints are evolutionarily conserved between mice and humans, however only indirectly. Comparative transcriptomic analysis of single drug and single gene perturbation experiments in mice and humans would be required to show this fact in a direct manner. Thus we conclude that it is reasonable to think that the footprints of a pathway or TF are evolutionarily conserved, at least at the level of our current footprint methods which rely on lists of genes.

Once shown that PROGENy and DoRothEA can also be applied to mouse data, we computed TF and pathway activities for a large collection of chemical and genetic perturbations and disease experiments. The results are provided as an interactive web application to browse corresponding pathway and TF activities (https://saezlab.shinyapps.io/footprint_scores). We envision that this resource can have broad applications including the study of diseases and therapeutics. Moreover, we demonstrated the usability of PROGENy and DoRothEA by recovering known pathway/TF disease associations using the aforementioned disease experiments. We found 4020 significant associations in total, where most were related to different forms of cancer, but we also recovered well-known associations of other disease types, such as liver disease (Fig. 4B).

Finally, we believe that our finding of the conserved nature of footprints is especially interesting for further development of footprint methods. Integrating data from mice and humans will provide a much stronger data background for future model construction. Lastly, we speculate that the conserved nature of footprints will not hold to be exclusively true for mouse and human but will also extend to other mammals often used as model organisms.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbgrm.2019.194431>.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

This work is supported by the German Federal Ministry of Education and Research (BMBF)-funded project Systems Medicine of the Liver (LiSyM, FKZ: 031L0049). B.S. was supported by the Premium Postdoctoral Fellowship Program of the Hungarian Academy of Sciences. We thank Luz Garcia-Alonso and Victoria Muckerson for discussion and useful feedback on the manuscript. We also thank Mi Yang and Ferenc Tajti for their help in the collection of perturbation experiments for the new PROGENy pathways.

Source code & data availability

All source code is deposited at <https://github.com/saezlab/ConservedFootprints>. Pathway and TF activities of perturbation and disease experiments can be browsed in a user friendly web application available at https://saezlab.shinyapps.io/footprint_scores.

Author contributions

C.H. and B.S. designed the research. C.H. performed all analyses, and wrote the manuscript. B.S. extended PROGENy with three new pathways. B.S. supported result interpretation and manuscript writing. J.S.R. supervised the project and contributed to writing the manuscript.

References

- [1] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550, <https://doi.org/10.1073/pnas.0506580102>.
- [2] M. Mann, O.N. Jensen, Proteomic analysis of post-translational modifications, *Nat. Biotechnol.* 21 (2003) 255–261, <https://doi.org/10.1038/nbt0303-255>.
- [3] M. Schubert, B. Klinger, M. Klünemann, A. Sieber, F. Uhlitz, S. Sauer, M.J. Garnett, N. Blüthgen, J. Saez-Rodriguez, Perturbation-response genes reveal signaling footprints in cancer gene expression, *Nat. Commun.* 9 (2018) 20, <https://doi.org/10.1038/s41467-017-02391-6>.
- [4] L. Cantini, L. Calzone, L. Martignetti, M. Rydenfelt, N. Blüthgen, E. Barillot, A. Zinoviyev, Classification of gene signatures for their information value and functional redundancy, *NPJ Syst Biol Appl.* 4 (2018) 2, <https://doi.org/10.1038/s41540-017-0038-8>.
- [5] The Mouse in Biomedical Research, Academic Press, 2007, <https://market.android.com/details?id=book-A237nQAACAAJ>.
- [6] R. Normand, W. Du, M. Briller, R. Gaujoux, E. Starosvetsky, A. Ziv-Kenet, G. Shalev-Malul, R.J. Tibshirani, S.S. Shen-Orr, Found In Translation: a machine learning model for mouse-to-human inference, *Nat. Methods* 15 (2018) 1067–1073, <https://doi.org/10.1038/s41592-018-0214-9>.
- [7] D.K. Brubaker, E.A. Proctor, K.M. Haigis, D.A. Lauffenburger, Computational translation of genomic responses from experimental model systems to humans, *PLoS Comput. Biol.* 15 (2019) e1006286, <https://doi.org/10.1371/journal.pcbi.1006286>.
- [8] J.D. Tenenbaum, M.G. Walker, P.J. Utz, A.J. Butte, Expression-based Pathway Signature Analysis (EPSA): mining publicly available microarray data for insight into human disease, *BMC Med. Genet.* 1 (2008) 51, <https://doi.org/10.1186/1755-8794-1-51>.
- [9] A. Aytes, A. Mitrofanova, C. Lefebvre, M.J. Alvarez, M. Castillo-Martin, T. Zheng, J.A. Eastham, A. Gopalan, K.J. Pienta, M.M. Shen, A. Califano, C. Abate-Shen, Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy, *Cancer Cell* 25 (2014) 638–651, <https://doi.org/10.1016/j.ccr.2014.03.017>.
- [10] L. Garcia-Alonso, C.H. Holland, M.M. Ibrahim, D. Turei, J. Saez-Rodriguez, Benchmark and integration of resources for the estimation of human transcription factor activities, *Genome Res.* 29 (2019) 1363–1375, <https://doi.org/10.1101/gr.240663.118>.
- [11] M.J. Alvarez, Y. Shen, F.M. Giorgi, A. Lachmann, B.B. Ding, B.H. Ye, A. Califano, Functional characterization of somatic mutations in cancer using network-based inference of protein activity, *Nat. Genet.* 48 (2016) 838–847, <https://doi.org/10.1038/ng.3593>.
- [12] Z. Wang, C.D. Monteiro, K.M. Jagodnik, N.F. Fernandez, G.W. Gundersen, A.D. Rouillard, S.L. Jenkins, A.S. Feldmann, K.S. Hu, M.G. McDermott, Q. Duan, N.R. Clark, M.R. Jones, Y. Kou, T. Goff, H. Woodland, F.M.R. Amaral, G.L. Szeto, O. Fuchs, S.M. Schüssler-Fiorenza Rose, S. Sharma, U. Schwartz, X.B. Bausela, M. Szymkiewicz, V. Maroulis, A. Salykin, C.M. Barra, C.D. Kruth, N.J. Bongio, V. Mathur, R.D. Todoric, U.E. Rubin, A. Malatras, C.T. Fulp, J.A. Galindo, R. Motiejunaite, C. Jüschke, P.C. Dishuck, K. Lahl, M. Jafari, S. Aibar, A. Zaravinos, L.H. Steenhuizen, L.R. Allison, P. Gamallo, F. de Andres Segura, T. Dae Devlin, V. Pérez-García, A. Ma'ayan, Extraction and analysis of signatures from the gene expression omnibus by the crowd, *Nat. Commun.* 7 (2016) 12846, <https://doi.org/10.1038/ncomms12846>.
- [13] A. Kauffmann, T.F. Rayner, H. Parkinson, M. Kapushesky, M. Lukk, A. Brazma, W. Huber, Importing ArrayExpress datasets into R/Bioconductor, *Bioinformatics* 25 (2009) 2092–2094, <https://doi.org/10.1093/bioinformatics/btp354>.
- [14] B. Phipson, S. Lee, I.J. Majewski, W.S. Alexander, G.K. Smyth, Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression, *Ann. Appl. Stat.* 10 (2016) 946–963, <https://doi.org/10.1214/16-AOS920>.
- [15] S. Durinck, P.T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt, *Nat. Protoc.* 4 (2009) 1184–1191, <https://doi.org/10.1038/nprot.2009.97>.
- [16] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinformatics* 12 (2011) 77, <https://doi.org/10.1186/1471-2105-12-77>.
- [17] J. Grau, I. Grosse, J. Keilwagen, PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R, *Bioinformatics* 31 (2015) 2595–2597, <https://doi.org/10.1093/bioinformatics/btv153>.
- [18] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, *Proceedings of the 23rd International Conference on Machine Learning*, ACM, New York, NY, USA, 2006, pp. 233–240, <https://doi.org/10.1145/1143844.1143874>.
- [19] S. Jupp, T. Burdett, C. Leroy, H.E. Parkinson, A new ontology lookup service at EMBL-EBI, *SWAT4LS*, 2015, pp. 118–119 http://ceur-ws.org/Vol-1546/paper_29.

- pdf.
- [20] A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation, *bioRxiv* (2016) 060012, <https://doi.org/10.1101/060012>.
 - [21] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, *Ann. Stat.* 29 (2001) 1165–1188 <http://www.jstor.org/stable/2674075>.
 - [22] E. Wingender, T. Schoeps, M. Haubrock, M. Krull, J. Dönitz, TFClass: expanding the classification of human transcription factors to their mammalian orthologs, *Nucleic Acids Res.* 46 (2018) D343–D347, <https://doi.org/10.1093/nar/gkx987>.
 - [23] H. Han, J.-W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C.Y. Kim, M. Lee, E. Kim, S. Lee, B. Kang, D. Jeong, Y. Kim, H.-N. Jeon, H. Jung, S. Nam, M. Chung, J.-H. Kim, I. Lee, TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions, *Nucleic Acids Res.* 46 (2018) D380–D386, <https://doi.org/10.1093/nar/gkx1013>.
 - [24] D.G. Johnson, The paradox of E2F1: oncogene and tumor suppressor gene, *Mol. Carcinog.* 27 (2000) 151–157 <https://www.ncbi.nlm.nih.gov/pubmed/10708476>.
 - [25] E. A. Fry, K. Inoue, Aberrant expression of ETS1 and ETS2 proteins in cancer, *Cancer Rep Rev.* 2 (2018), <https://doi.org/10.15761/CRR.1000151>.
 - [26] D.A. Fruman, C. Rommel, PI3K and cancer: lessons, challenges and opportunities, *Nat. Rev. Drug Discov.* 13 (2014) 140–156, <https://doi.org/10.1038/nrd4204>.
 - [27] M.I. Costache, M. Ioana, S. Iordache, D. Ene, C.A. Costache, A. Săftoiu, VEGF expression in pancreatic cancer and other malignancies: a review of the literature, *Rom. J. Intern. Med.* 53 (2015) 199–208, <https://doi.org/10.1515/rjim-2015-0027>.
 - [28] C.M. Quiñonez-Flores, S.A. González-Chávez, C. Pacheco-Tena, Hypoxia and its implications in rheumatoid arthritis, *J. Biomed. Sci.* 23 (2016) 62, <https://doi.org/10.1186/s12929-016-0281-0>.
 - [29] P.P. Tak, G.S. Firestein, NF-kappaB: a key role in inflammatory diseases, *J. Clin. Invest.* 107 (2001) 7–11, <https://doi.org/10.1172/JCI11830>.
 - [30] S. Banerjee, A. Biehl, M. Gadina, S. Hasni, D.M. Schwartz, JAK-STAT signaling as a target for inflammatory and autoimmune diseases: current and future prospects, *Drugs* 77 (2017) 521–546, <https://doi.org/10.1007/s40265-017-0701-9>.
 - [31] E.M. Zardi, L. Navarini, G. Sambataro, P. Piccinni, F.M. Sambataro, C. Spina, A. Dobrina, Hepatic PPARs: their role in liver physiology, fibrosis and treatment, *Curr. Med. Chem.* 20 (2013) 3370–3396 <https://www.ingentaconnect.com/content/ben/cmc/2013/00000020/00000027/art00004>.