The Signaling Pathways Project: an integrated 'omics knowledgebase for mammalian cellular signaling pathways

Scott Ochsner, David Abraham*, Kirt Martin*, Wei Ding, Apollo McOwiti, Zichen Wang, Kaitlyn Andreano, Ross A. Hamilton, Yue Chen, Angelica Hamilton, Marin L. Gantner, Michael Dehart, Shijing Qu, Susan G. Hilsenbeck, Lauren B. Becnel, Dave Bridges, Avi Ma'ayan, Janice M. Huss, Fabio Stossi, Charles E. Foulds, Anastasia Kralli, Donald P. McDonnell and Neil J. McKenna

Address Correspondence To:

Neil J. McKenna
Department of Molecular and Cellular Biology
Baylor College of Medicine
Houston, TX 77030
USA
e: nmckenna@bcm.edu
t: 713-798-8568

*These authors contributed equally to this study

BioRxiv DRAFT

Summary

Public transcriptomic and ChIP-Seq datasets have the potential to illuminate facets of transcriptional regulation by mammalian cellular signaling pathways not yet explored in the research literature. Unfortunately, a variety of obstacles prevent routine re-use of these datasets by bench biologists for hypothesis generation and data validation. Here, we designed a web knowledgebase, the Signaling Pathways Project (SPP), which incorporates stable community classifications of three major categories of cellular signaling pathway node (receptors, enzymes and transcription factors) and the bioactive small molecules (BSMs) known to modulate their functions. We then subjected over 10,000 publically archived transcriptomic or ChIP-Seq experiments to a biocuration pipeline that mapped them to their relevant signaling pathway node, BSM or biosample (tissue or cell line of study). To provide for prediction of pathway node-target transcriptional regulatory relationships, we generated consensus 'omics signatures, or consensomes, based on the significant differential expression or promoter occupancy of genomic targets across all underlying transcriptomic (expression array and RNA-Seq) or ChIP-Seq experiments. To expose the SPP knowledgebase to biology researchers, we designed a web browser interface that accommodates a variety of routine data mining strategies depending upon the requirements of the end user. Individual dataset pages provide for browsing or filtering, and facilitate integration of SPP with the research literature. Results of single gene, Gene Ontology or user-uploaded gene list queries are displayed in an interactive user interface referred to as the Regulation Report, in which evidence for transcriptional regulation of downstream genomic target by cellular signaling pathway nodes is compartmentalized in an intuitive manner. Consensome queries allow users to evaluate evidence for targets most consistently regulated by a given signaling pathway node family, and allow for detailed inspection of the pharmacology underlying node-target regulatory relationships predicted by the consensomes. Consensomes were validated using alignment with literature-based knowledge,

BioRxiv DRAFT

gene target-level integration of transcriptomic and ChIP-Seq data points, and in bench experiments that confirmed previously uncharacterized node-gene target regulatory relationships. SPP is freely accessible at https://beta.signalingpathways.org.

BioRxiv DRAFT

Availability and Implementation: The Signaling Pathways Project is freely accessible at https://beta.signalingpathways.org.

Social media: @sigpathproject

BioRxiv DRAFT

Introduction

Signaling pathways describe functional interdependencies between distinct classes of molecules that collectively determine the response of a given cell to its afferent metabolic and endocrine signals [1]. The bulk of readily accessible information on these pathways resides in the conventional research literature in the form of peer-reviewed hypothesis-driven research articles, and in knowledgebases that curate such information [2]. Many such articles are based in part upon discovery-scale datasets documenting, for example the effects of genetic or small molecule perturbations on gene expression in transcriptomic datasets, and DNA promoter region occupancy in cistromics, or ChIP-Seq, datasets. Conventionally, only a small fraction of data points from such datasets are characterized in any level of detail in associated hypothesis-driven articles. While largely unused initially, the remaining data points in 'omics datasets possess potential collective re-use value for validating experimental data or gathering evidence to model cellular signaling pathways. We and others have described the limited findability and accessibility, interoperability and re-use (FAIR) status of these datasets [3, 4]. Although some barriers to the FAIR status of these datasets are being addressed, and a number of useful' omics dataset-based research resources have been developed [5-14], opportunities exist to further develop the infrastructure enabling routine re-use of public 'omics datasets by bench researchers in the field of mammalian cellular signaling.

We previously described biocuration and web development approaches to enhance the FAIR status of public transcriptomic datasets involving genetic or small molecule perturbations of members of the nuclear receptor (NR) superfamily of ligand-regulated transcription factors [15]. Here we describe a distinct and original knowledgebase, the Signaling Pathways Project (SPP), which expands these FAIR efforts along three dimensions. Firstly, we have encompassed datasets involving genetic and small molecule perturbations of a broad range of cellular signaling pathway modules - receptors, enzymes, transcription factors. Secondly, we have

BioRxiv DRAFT

integrated ChIP-Seq datasets, which document genomic occupancy by transcription factors, enzymes and other factors. Thirdly, we have developed a meta-analysis technique that surveys across these datasets to generate consensus ranked signatures, referred to as consensomes, which allow for prediction of signaling pathway node-target regulatory relationships. We validate the consensomes using alignment with literature knowledge, integration of transcriptomic and ChIP-Seq evidence, and using bench experimental use cases that validate signaling pathway node-target regulatory relationships predicted by the consensomes. Finally, we have made the entire data matrix available for routine data browsing, mining and hypothesis generation by the mammalian cell signaling research community at https://beta.signalingpathways.org.

BioRxiv DRAFT

Results

Data model design

The goal of the Signaling Pathways Project (SPP) is to give bench scientists routine access to biocurated public transcriptomic and ChIP-Seq datasets to infer or validate cellular signaling pathways operating within their biological system of interest. Although such pathways are diverse and dynamic in nature, they typically describe functional interdependencies between molecules belonging to three major categories of pathway module: activated transmembrane or intracellular receptors, which initiate the signals; intracellular enzymes, which propagate and modulate the signals; and transcription factors, which give effect to the signals through regulation of gene expression [16]. Accordingly, we first set out to design a knowledgebase that would reflect this modular architecture. To ensure that our efforts were broadly aligned with established community standards, we started by integrating existing, mature classifications for receptors (International Union of Pharmacology, IUPHAR; [17]), enzymes (International Union of Biochemistry and Molecular Biology Enzyme Committee [18]) and transcription factors (TFClass [19]). Table S1 shows representative examples of the hierarchical relationships within each of the signaling pathway module categories. To harmonize and facilitate data mining across different signaling pathway modules, top level categories were subdivided firstly into functional classes, which in turn were subdivided into gene families, to which individual gene products were assigned. Fig. 1 summarizes the major classes and families in each category encoded in the data model. Consistent with terminology in use in the cellular signaling field [1, 20], we refer to these individual gene products as nodes. Molecular classes that are relevant to, but less frequently studied in the context of cellular signaling, such as regulatory RNAs, chromatin factors and cytoskeletal components, were assigned to a Co-nodes category. Impacting the functions of nodes in all four categories are bioactive small molecules (BSMs), encompassing:

BioRxiv DRAFT

physiological ligands for receptors; prescription drugs, targeting almost exclusively nodes in the receptor and enzyme categories; synthetic organics, representing experimental compounds and environmental toxicants; and natural products (S1 Table). BSM-node mappings were retrieved from an existing pharmacology biocuration initiative, the IUPHAR Guide To Pharmacology [17], or annotated by SPP biocurators *de novo* with reference to a specific PubMed identifier (PMID).

Dataset biocuration

Having defined relationships within each major signaling pathway module, we next designed a dataset biocuration strategy that would classify publically archived transcriptomic and ChIP-Seq datasets according to the signaling pathway node(s) whose transcriptional functions they were designed to interrogate (Fig. S1). For knowledgebase design purposes, we defined a dataset as a collection of individual experiments encompassed by a specific GEO series (GSE, for transcriptomic datasets) or SRA Project (SRP, for ChIP-Seq datasets).

Transcriptomic datasets

We previously described our efforts to biocurate Gene Expression Omnibus (GEO) transcriptomic datasets pertinent to nuclear receptor signaling as part of the Nuclear Receptor Signaling Atlas [15]. In order to expand this collection to encompass datasets involving perturbation of a broader range of signaling pathway nodes, we carried out a systematic survey of Gene Expression Omnibus to identify an initial population of transcriptomic datasets constituting a representative cross-section of the various classes of signaling pathway node referred to in Fig. 1. To supplement this effort, we also incorporated datasets from the CREEDS project [21], a crowd-based initiative that systematically identified and annotated single gene and BSM perturbation GEO datasets. From this initial collection of datasets, we next carried out a three step QC check to filter for datasets that (i) included all files required to calculate gene differential expression values; (ii) contained biological replicates to allow for calculation of

BioRxiv DRAFT

associated significance values; and (iii) whose samples clustered appropriately by principal component analysis. Typically, 20-25% of archived transcriptomic datasets were discarded at this step. The remaining datasets were diverse in design, typically involving genetic (single or multi-node node overexpression, knockdown, knockin or knockout) or BSM (physiological ligand, drug or synthetic organic or natural substance; single or multi-BSM; time course; agonist, antagonist or tissue-selective modulator) manipulation of a signaling node across a broad range of human, mouse and rat biosamples. To maximize the amount of biological information extracted from each transcriptomic dataset, we calculated differential expression values for all possible contrasts, and not just those used by the original investigators in their publications. Next, transcriptomic experiments were mapped where appropriate to approved symbols (AGSs) for human, mouse and rat genes, representing genetically perturbed signaling nodes, and/or to unique identifiers for BSMs, as well as to a previously described biosample controlled vocabulary. Gene differential expression values were calculated for each experiment using an industry standard Bioconductor pipeline [15]. Finally, experiments were organized into datasets for which digital object identifiers (DOIs) were minted as previously described [4].

ChIP-Seq datasets

In addition to integration of transcriptomic datasets with each other, their integration with related ChIP-Seq datasets was desirable since it would provide for cross validation of predicted node-target relationships, as well as providing for more detailed mechanistic modeling of such relationships than would be possible using either omics platform individually. The ChIP-Atlas resource [22] supports re-use of ChIP-Seq datasets by carrying out uniform MACS2 peak-calling across ChIP-Seq datasets archived in NCBI's Short Read Archive (SRA). We therefore next set out to identify and incorporate ChIP-Atlas-processed SRA ChIP-Seq datasets relevant to mammalian signaling pathway nodes. Individual SRA experiments were first mapped to the AGS of the immunoprecipitation (IP) node and any other genetically manipulated nodes (e.g.

BioRxiv DRAFT

knockdown or knockout background), to any BSMs represented in the experimental design, and to the biosample in which the experiment was carried out.

Generation of consensomes

An ongoing challenge for the cellular signaling bioinformatics research community is the meaningful integration of the universe of ‘omics data points to enable researchers lacking computational expertise to develop focused research hypotheses in a routine and efficient manner. A particularly desirable goal is unbiased meta-analysis to define community consensus reference signatures that allow users to predict regulatory relationships between signaling pathway nodes and their downstream targets. Accordingly, we next set out to design a meta-analysis pipeline that would leverage our biocurational platform to reliably identify signaling pathway node - target gene regulatory relationships in a given biosample context. Since this analysis was designed to establish a consensus across distinct datasets from different laboratories, we referred to it as consensomic analysis, and the resulting node-target rankings as consensomes.

Transcriptomic consensomes

Large scale meta-analysis pipeline of publically archived transcriptomic datasets is confronted primarily by the sheer heterogeneity of genetic and pharmacological perturbation designs represented in these datasets. We hypothesized that irrespective of the nature of the perturbation impacting a given pathway node, downstream targets with a greater dependence on the integrity of that node would be more likely to be differentially expressed in response to its perturbation than those with a weaker regulatory relationship with the node. Accordingly, to enhance the statistical power of the analysis, we initially binned transcriptomic experiments for meta-analysis on the basis of genetic or pharmacological manipulation of a given signaling node. To further extend statistical power, experiments involving manipulation of all nodes in a

BioRxiv DRAFT

defined gene family were combined for meta-analysis. Next, we further classified experiments according to the biosample and species in which they were carried out. Gene target-specific nominal p-values and differential expression values were then aggregated over each defined set of experiments to yield target-specific summaries for a consensome. A more detailed description of the transcriptomic consensome algorithm is contained in File S1.

A number of factors determine whether a target will be induced or repressed by manipulation of a given signaling node in any given experiment. These include: node isoform differential expression [23]; cell cycle stage [24]; biosample of study [25]; BSM dose treatment duration; and perturbation type (loss or gain of function). To avoid these opposing alterations canceling each other out at the target transcript level in the meta-analysis, we converted fold changes to positive fold changes (i.e. max (FC, 1/FC)) so that both induction and repression would be counted as 'altered' in a summary measure of the magnitude of perturbation, which was computed as the geometric mean fold change. In addition, for each target, we counted the number of experiments with gene-specific nominal p-values ≤0.05, and computed the binomial probability, referred to as the consensome p-value (CPV), of observing that many or more nominally significant experiments out of the number of experiments in which the target was assayed, given a true probability of 0.05. Targets were then ranked in consensomes in ascending order of the consensome p-value (CPV), with average rank being reported for tied CPVs.

ChIP-Seq consensomes

For calculation of ChIP-Seq consensomes, groups of experiments were formed whose IP nodes mapped to a defined node family. These classes were further sorted into meta-analysis classes based on mapping to the same biosample controlled vocabulary used to annotate the transcriptomic datasets [15]. In contrast to the transcriptomic consensomes, which were based

BioRxiv DRAFT

upon differential expression (DE) and significance values generated *de novo* from raw files, MACS2 peak calls and associated significance cut-offs were retrieved in pre-processed form from ChIP-Atlas [22].

Signaling Pathways Project user interface

To make the results of our biocuration and analysis routinely and freely available to the research community, we next developed a web interface for the SPP knowledgebase that would provide for browsing of datasets, as well as for mining of the underlying data points. A comprehensive walkthrough file containing instructions on the use of the SPP interface is shown in File S2.

Browsing of SPP datasets

The full list of SPP datasets can be filtered using any combination of 'omics type, signaling pathway category, class or family, biosample physiological system and organ, or species. Individual dataset pages enable integration of SPP with the research literature via DOI-driven links from external sites, as well as for citation of datasets to enhance their FAIR status [3, 4]. To accommodate users seeking a rapid summary of the targets with the highest differential expression (for transcriptomic datasets, example: analysis of the sperm-specific antigen 2 (Ssfa2)-dependent transcriptome in mouse liver), or highest MACS2 peak value (for ChIP-Seq datasets, example: analysis of the CREBBP cistrome in human embryonic kidney 293 cells), the most highly ranked targets in a given experiment are displayed. The user can toggle between individual experiments using a pull-down menu.

Mining of SPP datasets in Ominer

The SPP query interface, Ominer, allows a user to specify single gene target, GO term or a custom gene list in the "Gene(s) of Interest" drop-down, and to dial in additional node and biosample regulatory parameters in subsequent drop-down menus as required (Fig. 2A).

BIORXIV DRAFT

Examples of single gene and GO term queries are shown in Table 1 and Table 2, respectively. Results are returned in an interface referred to as the Regulation Report, a detailed graphical summary of evidence for transcriptional regulatory relationships between signaling pathway nodes and a genomic target(s) of interest (Fig. 2, B & C). To allow users to share links to SPP Regulation Reports with colleagues, or to embed them in research manuscripts or grant applications, the Reports are accessed by a constructed URL defining all of the individual query parameters. The vertical organization of the default Category view in both Regulation Reports reflects conventional schematic depictions of cellular signaling pathways, with Receptors on top, followed by Enzymes, then Transcription Factors (Fig. 2, B & C). A fourth category, Co-nodes, contains factors that are less frequently studied in the context of cellular signaling pathways. Consistent with the hierarchy in Table S1, each Regulation Report category is subdivided into classes (depicted as **Category | Class** in the UI, Fig. 2, B & C) which are in turn subdivided into families, which in turn contain member nodes, which are themselves mapped to BSMs (Fig. 2, B & C). The transcriptomic Regulation Report displays differential expression levels of a given target in experiments involving genetic (rows labelled with italicized node AGS) or BSM (rows labelled with bold BSM symbol) BSM manipulations of nodes within a given family (Fig. 2B). The cistromics/ChIP-Seq Report displays MACS2 peak values within 10 kb of a given promoter transcriptional start site (TSS) in ChIP-Seq experiments named using the convention IP Node AGS | **BSM Symbol** | *Other Node AGS* (Fig. 2C).

To accommodate users seeking a perspective on regulation of a target in a specific organ, tissue, cell line or species, users can select the “Biosample” and “Species” views from the dropdown (Fig. 2B). Each data point in either Regulation Report links to a pop-up window containing the essential experimental information (Fig. 2D, upper = transcriptomic, lower = cistromic). This in turn links to a window summarizing the pharmacology of any BSMs used in the experiment (Fig. 2E), or a Fold Change Details window that places the experiment in the

BioRxiv DRAFT

context of the parent dataset (Fig. 2F), linking to the full dataset page and associated journal article. The Fold Change Details window also provides for citation of the dataset, an important element of enhancing the FAIR status of ‘omics datasets [3].

Consensomes: discovering downstream transcriptional targets of signaling pathway nodes

Table 3 shows examples of the consensomes available in the initial version of the SPP knowledgebase. Consensomes can be accessed through Ominer, in which the user selects the “Consensome” from “Genes of Interest”, then either “Transcriptomic” or “Cistromic (ChIP-Seq)” from the “’Omics Category” menu. Subsequent menus allow for selection of specific signaling pathway classes or families, physiological system or organ of interest, or species. To accommodate researchers interested in a specific physiological system or organ rather than a specific signaling node, consensomes are also calculated across all experiments mapping to a given physiological system (metabolic, skeletal) and organ (liver, adipose), providing for identification of targets under the control of a broad spectrum of signaling nodes in those organs (Table 3). To maximize their distribution and exposure in third party resources, consensomes can also be accessed by direct DOI-driven links.

Consensomes are displayed in an accessible tabular format in which the default ranking is in ascending order of CPV, although targets can be ranked by any column desired (Fig. 3). To reflect the frequency of differential expression of a target relative to others in a given consensome, the percentile ranking of each target within the consensome is displayed. Targets in the 90th percentile of a given consensome – the highest confidence predicted targets for a given node family - are accessible through the web interface, and the entire list of targets is available for download in spreadsheet format for import into custom analysis programs. As previously discussed, to suppress the diversity of experimental designs as a confounding

BioRxiv DRAFT

variable in consensome analysis, the direction of differential expression is omitted when calculating the ranked signatures. An appreciation of the pharmacology of a specific node-target gene relationship is essential however to allow researchers to place the ranking in a specific biological context and to design subsequent experiments in an informed manner. To accommodate this, the target gene symbols in consensomes link to transcriptomic or cistromic Regulation Reports filtered by family and/or biosample to display those data points that contributed to the calculation of the specific consensome.

A useful feature of the consensome table is the ability to filter the list by target gene symbol using the Search box (Fig. 3). Although this can be used for identifying a single gene of interest, it also illuminates potentially biologically significant regulation of targets encoding multiple members of a gene family by a given node. For example the significant enrichment in the 90th percentile of the human estrogen receptor family (ERs-Hs-All systems) consensome of multiple members of the go-ichi-ni-san (HGNC root symbol GINS; [26]), condensin (NCAP; [27]), minichromosome maintenance (MCM; [28]) and centromere protein (CENP; [29]) families, among others reflects the profound impact of estrogen receptor signaling on DNA replication and cell division in its target organs.

Validation of consensomes

The design of the transcriptomic consensome analysis was predicated upon three assumptions: firstly, that borrowing statistical power by binning experiments according to their perturbation of a given signaling node was biologically valid; secondly, that omitting direction of differential expression from the analysis would allow for direct interrogation of the strength of the regulatory relationship between a node and a target, independent of the nature of the node perturbation used in an experiment; and thirdly, that ranking targets according to the frequency of their significant differential expression, rather than by fold change, accurately reflected the relative

BioRxiv DRAFT

strengths of the regulatory relationship between a given node and its transcriptional targets. We next wished to determine whether these assumptions were legitimate, and to establish whether the consensomes were indeed reliable consensus regulatory signatures for cellular signaling nodes. We designed a consensome validation strategy comprising four components: comparison of consensomes with existing canonical (i.e. literature-based) node-target relationships; reciprocal validation of node-target relationships between transcriptomic and ChIP-Seq consensomes; cross-validation between distinct family & node consensomes within a single organ; and experimental bench validation of predicted node-target relationships.

Three considerations recommended members of the nuclear receptor (NR) superfamily of physiological ligand-regulated transcription factors for selection for proof-of-principle validation of the consensomes. Firstly, as the largest single class of drug targets, they are the subject of a large body of dedicated research literature, affording considerable opportunity for testing the consensomes against existing canonical knowledge. Secondly, as ligand-regulated transcription factors, members of this superfamily are prominently represented in both publically archived transcriptomic and ChIP-Seq experiments, enabling meaningful cross-validation of consensomes between these two experimental categories. Finally, by combining two pathway node categories (receptor and transcription factor) in a single molecule, they allow for validation of two node functions in a single bench experiment.

Canonical signaling node targets are highly ranked in consensomes

To compare consensome rankings with canonical node-target relationships, we selected the ten top ranked targets in the ER subfamily in human mammary gland (ERs-Hs-MG), the androgen receptor in human prostate gland (AR-Hs-Prostate), the glucocorticoid receptor in the human metabolic system (GR-Hs-Metabolic), and the peroxisome proliferator-activated receptor (PPAR) family in the mouse metabolic system (PPARs-Mm-Metabolic). We then searched the

BioRxiv DRAFT

research literature to identify articles in which these genes had been functionally characterized as targets of these receptors. As shown in Table S2, the most highly ranked targets for the AR-Hs-All, ER-Hs-All and PPARs-Mm-Metabolic consensomes were well supported by evidence in the research literature, although overlap with literature knowledge was lower for the GR-Hs-Metabolic consensome (Table S2).

Reciprocal validation of transcriptomic and ChIP-Seq consensomes

Although many of the most highly ranked consensome target were validated by prior characterization in the research literature some, such as those in the GR-Hs-Metabolic consensome, were not. What was unclear at this juncture was whether such genes were authentic target genes, and therefore represented gaps in literature knowledge that were filled by the consensomes, or they were false positives, and their elevated consensome rankings were therefore misleading. To distinguish between these two possibilities, we next wished to determine the extent to which NR node-target relationships predicted by the transcriptomic consensomes were validated by the publically archived ChIP-Seq datasets involving the corresponding receptors. In the canonical model of NR signaling binding of endogenous ligands such as 17β-estradiol (17BE2) or dihydrotestosterone (DHT), NRs are released from inhibitory heat shock proteins, spontaneously dimerize and translocate to the nucleus where they interact with specific promoter enhancer elements to regulate expression of target genes [30]. Of the 40 genes selected for literature validation (Table S2), 90% (45/50) were in the 90th percentile or higher in the corresponding ChIP-Seq consensomes, indicating that they are regulated at least in part by direct receptor-enhancer interactions. Interestingly, of the eight transcriptomic consensome-predicted node-target relationships for which no supporting literature evidence was found, all but one were in the 90th percentile or higher of the corresponding ChIP-Seq consensome.

BioRxiv DRAFT

Intersections of transcriptomic consensomes for key hepatic signaling nodes are enriched for targets encoding critical metabolic pathway enzymes

Transcriptional regulation of metabolism by cellular signaling pathways is a well-established paradigm [31]. Consistent with this, a broad range of hepatic pathways impacting metabolism of carbohydrate, lipids, amino acids and other intermediates are under fine transcriptional regulation by a variety of nuclear receptors, including NR1H4/FXR [32], NR3C1/GR [33] and members of the PPAR [34, 35] families. If our assertion that consensomes reflected the relative strengths of node-target regulatory relationships was valid, we anticipated that gene targets with elevated rankings across these three hepatic consensomes (and, by implication, strong regulatory relationships with these receptors) would be enriched for targets encoding factors with prominent roles in hepatic metabolism. To test this hypothesis, we first identified genes in the All nodes-Mm-liver TC90 (n = 1999), that is, those genes that were in the top 10% of targets that are differentially expressed in expression profiling experiments in a murine hepatic biosample, irrespective of the perturbed signaling node. This gene set was then filtered for targets that were present in the 90th percentile of all three of the PPARs-Mm-liver, GR-Mm-liver and FXR-Mm-liver TC90s (n = 117; Fig. 5A and File S3). Combining a curated database of metabolic enzymes ([36] n = 1647) with a literature search, we found that 39 (33%) of the intersection targets encoded metabolic enzymes (Fig. 4A and File S3). This represented a 4-fold enrichment of such genes compared with 8/95 (8%) in a random sample of the original full All nodes-Mm-liver consensome (n=20,000), a proportion comparable to that estimated (7%) for metabolic enzymes across the entire protein-encoded genome [36]. Of these 39 enzymes, a literature search showed that nearly two-thirds (23/36) regulate a rate limiting or committed step in a metabolic pathway (Table 4 and bold in Fig. 4B) and/or are deficient or mutated in a known metabolic disorder (Table 4 and marked with an asterisk* in Fig. 4B). Many of these enzymes are historically well characterized, including Acaca, which regulates the rate limiting step in fatty

BioRxiv DRAFT

acid synthesis [37] and is deficient in acetyl CoA carboxylase syndrome [38], and Hal, which regulates the initial step in histidine catabolism and is deficient in hisitidinemia [39]. The critical metabolic roles of other enzymes however, such as Nnmt [40] and Parp14 [41], have been only much more recently characterized. In addition to enzymes, other nodes that participate in pathways with critical roles in hepatocyte homeostasis and development, such as Il6ra ([42], Cebpb [43] and members of the Irf transcription family [44] are represented at the intersection of the four consensomes (File S3). This analysis demonstrates the ability of organ level consensomes to illuminate factors that are downstream targets of multiple signaling nodes and, by extension, have pivotal, tightly-regulated roles in the function of a given physiological system or organ.

Bench validation: elevated consensome rankings predict biological node-target relationships

A primary motivation in developing the SPP resource was to assist researchers in filling gaps in the literature regarding knowledge of cellular signaling pathways. Indeed, in addition to corroborating canonical node-target gene relationships, we found that the node transcriptomic consensomes contained targets that had elevated percentile rankings, but were uncharacterized in the research literature with respect to regulation by that node. Accordingly, we next set out to experimentally validate representative examples of these targets, shown in Table S3, at the bench.

TPD52L1 is a stress fiber-associated factor that supports 17BE2-dependent MCF-7 cell proliferation

We first wished to broadly evaluate the extent to which experimental evidence validated the node-target relationships predicted by the consensomes. To do this, we used Q-PCR to evaluate 17BE2-dependent regulation of a panel of both characterized and uncharacterized ER

BioRxiv DRAFT

targets that were highly ranked in the ER-Hs consensome (Fig. 5A). Reflecting their elevated consensome rankings, the expression of all the genes tested were found to be regulated by 17BE2 in either a dose dependent (*GREB1*, *TPD52L1* and others), or a biphasic (*MYC* and *TFF1*) manner, activated and suppressed at physiological or supraphysiological levels of 17BE2, respectively. We next wanted to evaluate the dependence of this regulation on the integrity of nodes in the ER family (ESR1 and ESR2) using the selective ER downregulator fulvestrant (FULV), which blocks the function of these nodes by disrupting their interaction with 17BE2 and inducing their proteasomal degradation [45]. Consistent with the strong ER family node dependence of their regulation predicted by the ER-Hs-mammary gland transcriptomic and ChIP-Seq consensomes, FULV completely abolished 17BE2 induction of all target genes tested (Fig. 5A).

We next selected one of the uncharacterized ER consensome targets for further study. The tumor protein D52-like 1 (*TPD52L1*) gene encodes a little-studied protein that bears sequence homology to members of the TPD52 family of coiled-coil motif proteins that are overexpressed in a variety of cancers [46]. Despite a ranking in the transcriptomic (ERs-Hs-All-TC CPV = 1E-130, 99th percentile) and ChIP-Seq (ERs-Hs-All-CC 99th percentile) ER consensomes that was comparable to or exceeded that of canonical ER target genes such as GREB1 or MYC, and subsequent experimental bench validation of the ER family-*TPD52L1* regulatory relationship (Fig. 5A), no evidence for regulation of ER by TPD52L1 was found in the research literature. Interestingly, peak cell cycle expression of both *TPD52L1* [47] and *ESR1* occur at the G2-M transition, which is also a point at which ESR1 is known to exert control of cell cycle progression [48]. Based upon these observations, we selected *TPD52L1* for further validation and characterization in the context of ER signaling. Immunofluorescene analysis of TPD52L1 in MCF-7 cells demonstrated specific 17BE2-dependent association of TPD52L1 with numerous structures, including nucleus, plasma membrane, cytoplasm and stress fiber-like structures (Fig.

BioRxiv DRAFT

5B), which play an important role in mitosis orientation, a critical process in cell division [49]. Having established a potential function for TPD52L1 in regulation of the cell cycle, we hypothesized that its depletion in cells might block this function and retard cell growth. Consistent with this, and in support of previously-reported associations of its family member *TPD52* with increased proliferation and invasive capacity [50, 51], we found that siRNA-mediated knockdown of *TPD52L1* by 80% (data not shown) resulted in a significant decrease in 17BE2-induced proliferation of MCF-7 cells (Fig. 5C). Interestingly, the *TPD52L1* transcriptomic Regulation Report showed disruption of *TPD52L1* expression in response to manipulation of kinases in the checkpoint (CHEK1, CHEK2), cyclin-dependent kinase (CDK9) and MAPK superfamily (ATR, ATM and RAF1) are consistent with known roles for these enzymes in regulation of the G2/M checkpoint [52-56] . In concert, these observations constitute experimental validation of the biological relationship between ER signaling and TP52L1 predicted by the ER family-Hs-mammary gland transcriptomic and ChIP-Seq consensomes.

MBOAT2 connects phospholipid metabolism to AR regulation of prostate cell growth

The next bench validation use case illustrates the value of integration of ChIP-Seq data points in gathering evidence to establish the plausibility of a node-target relationship implied by the consensomes. The *MBOAT2* gene encodes an enzyme, membrane-bound O-acyl transferase 2, that catalyzes cycles of glycerophospholipid deacylation and reacylation to modulate plasma membrane phospholipid asymmetry and diversity [57]. We noted that the ranking of MBOAT2 in both the AR-Hs-All TC (CPV = 2.2E-35, 99th percentile) and ChIP-Seq (99th percentile) consensomes was comparable to that of the canonic and intensively studied AR target genes such as *KLK3* and *TMPRSS2*. In contrast to the large volume of literature these targets however, with the exception of a mention in a couple of androgen expression profiling studies [58, 59], the role of MBOAT2 in the context of AR signaling was entirely unstudied. Our attention was drawn to *MBOAT2* as a candidate for bench validation as an AR downstream target by a

BioRxiv DRAFT

number of different lines of evidence. Firstly, in addition to numerous AR binding data points, the *MBOAT2* cistromic/ChIP-Seq Regulation Report contained evidence for binding sites within 10 kb of the *MBOAT2* TSS for the transcription factors GATA1, FOXA1, MYC and NANOG, all of which have known roles in AR crosstalk [60]. Encouraged by the cistromic/ChIP-Seq evidence corroborating its elevated ranking in the AR-Hs-All TC, we selected *MBOAT2* for further validation and characterization. We first wished to test whether *MBOAT2* was an AR-regulated gene in cultured prostate cancer cell lines. As shown in Fig. 5D, *MBOAT2* was induced in LNCaP prostate epithelial cells in response to treatment with the physiological AR agonist dihydrotestosterone (DHT). We next determined the effect of depletion of MBOAT2 on LNCaP cell viability and found that relative to control siRNA treatment, siMBOAT2 significantly increased LNCaP cell numbers at growth day 5 in in R1881-treated cells, but not untreated cells (Fig. 5E).

This result was unexpected to us given the prevailing perception of AR as a driver of prostate tumor growth, but can be rationalized in the context of suppression of growth and support of differentiation by AR in normal prostate luminal epithelium [61]. This process is known to involve induction of the NKX3.1 (AGS: *ZBTB16*) homeobox transcription factor [62, 63] - itself the highest ranked gene in the AR-Hs-All transcriptomic consensome - and it can be speculated that induction of *MBOAT2* by AR represents an additional component of this process. Such an assertion is supported by the recent characterization of the role of *MBOAT2* in chondrogenic differentiation of ATDC5 cells [64], and by the fact that the AR agonist testosterone stimulates the chondrogenic potential of chondrogenic progenitor cells [65].

GR and ERR exert co-ordinate regulation of glycogen metabolism via regulation of protein phosphatase subunit expression

BioRxiv DRAFT

The first two experimental validation studies focused on distinct single node-target regulatory relationships. We next wished to validate the use of consensome intersection analysis to highlight convergence of multiple signaling nodes on targets involved in a common downstream biological process. Interconversion of glucose and glycogen in metabolic organs is under the tandem control of glycogen synthase and glycogen phosphorylase, which respectively promote and restrict the incorporation of glucose into glycogen in response to hormonal and stress regulatory cues [66]. The activity of glycogen synthase is in turn under the control of protein phosphatase 1 (PP1), which converts it from its inactive phosphorylated form to its active dephosphorylated form [67], and 5'AMP-activated protein kinase (AMPK), which catalyzes the reverse reaction [68] (Fig. 6A). Although historical evidence indicates that glucocorticoids promote glycogen storage in the liver through upregulation of glycogen synthase phosphatase activity [69], the underlying mechanism has to date been unclear. Similarly, although members of the ERR subfamily have been shown to promote reprogramming of carbohydrate metabolism in exercising skeletal muscle [70], a direct role for ERRs in controlling glycogen turnover in muscle had not been investigated. Two key regulatory subunit genes relevant to PP1 and AMPK are *Ppp1r3c*, encoding PTG in the PP1 holoenzyme [71, 72], and *Prkab2*, encoding AMPKβ2 in the AMPK holoenzyme [73, 74]. Based on the significant rankings of *PP1R3C* in the GR human (CPV = 3.2E-08) and mouse (CPV = 3.4E-10) transcriptomic consensomes, and both *PPP1R3C* (CPV = 2.7E-06) and *PRKAB2* (E=6.3E-05) in the ERR-Hs-All transcriptomic consensome, we hypothesized that the mechanism by which these receptors controlled carbohydrate metabolism in liver and skeletal muscle might encompass regulation of expression of these two genomic targets.

Based upon Ominer Regulation Report evidence for binding of GR to the *Ppp1r3c* promoter in mouse liver, we undertook sequence analysis of the murine *Ppp1r3c* promoter and identified two prominent peaks 5' to the first exon of *Ppp1r3c*, the more proximal peak of which contained

BioRxiv DRAFT

a potential glucocorticoid response element (GRE, Fig. 6B; based on the GRE consensus [75]). To determine whether *Ppp1r3c* was upregulated in isolated cells, we treated a hepatoma cell line with the synthetic glucocorticoid dexamethasone (DEX) for 48 h and observed upregulation of *Ppp1r3c* mRNA (Fig. 6C). As positive controls, we also noted observed hepatic induction of the genes encoding pyruvate carboxylase (*Pcx*) [76] and *Fgf21* [77], established GR targets that have significant rankings in the GR-Mm-All transcriptomic consensome.

We next wished to determine whether the same *Ppp1r3c* PP1 regulatory subunit gene targeted by GR, as well as the AMPK subunit gene *Prkab2*, were directly regulated by ERRs. Evidence in the SPP cistromic Regulation Reports for Ppp1r3c and Prkab2 and from IVG analysis of additional datasets (Fig. 6D) supported the presence of one or more Esrra binding sites within 10 kb of the *Ppp1r3c* and *Prkab2* TSSs. To investigate the effect of small molecule manipulation of Esrra on endogenous expression of *Prkab2* or *Ppp1r3c*, we treated C2C12 myotubes (day 3 MT) for 24 h with BSM inhibitors of Esrra (Fig. 6E). *Prkab2* was repressed by the Esrra inverse agonist XCT790 [78] (Fig. 5E, right panel) whereas *Ppp1r3c* transcript expression was unaffected in response to this treatment (Fig. 5E, left panel). We next evaluated the effects on endogenous *Ppp1r3c* and *Prkab2* expression of genetic manipulation of ERR signaling using adenoviral overexpression of Esrra (Fig. 6F). Interestingly, whereas *Ppp1r3c* was upregulated in response to Esrra gain of function (Fig. 6F, left panel), expression of *Prkab2* was not significantly impacted (Fig. 6F, right panel). The differential regulation of the two targets in these experiments suggests that Esrra may be more important for maintaining basal expression of *Prkab2* and mediating regulation of *Ppp1r3c* expression in response to physiologic stimuli. We next assessed whether the expression of Prkab2 was altered in Esrra-deficient skeletal muscle [79]. Consistent with its elevated ERR consensome ranking, basal expression of *Prkab2* transcript was reduced by 40% in Esrra-depleted skeletal muscle compared to wild-type tissue (Fig. 6G).

BioRxiv DRAFT

To test if ERRs directly regulate the *Prkab2* target, the -2820 to +27 region (relative to the TSS +1) was cloned upstream of a luciferase reporter gene. Based on JASPAR transcription factor binding site prediction software [80], we determined that this region contained a number of high scoring ERR family consensus binding sites (data not shown). Several of the predicted ERRE sites were in close proximity to consensus sites for Gabpa, Creb, and Stat3, which are often clustered with ERREs and are known to facilitate functional interactions between ERR family members and these transcription factors [81]. Consistent with this, the *PRKAB2* ChIP-Seq Regulation Report contains evidence for binding of CREB and Stat factors to the *PRKAB2* promoter. In transcriptional assays performed in C2C12 myoblasts we observed a similar magnitude of activation of the *Prkab2* promoter-reporter in response to co-transfected Esrra or Esrrg (Fig. 6H). We then assessed whether the regulation of *Prkab2* by Esrr was impacted by insulin-like growth factor 1 (IGF1), which signals through AKT and MAPK to promote myocyte glucose uptake and glycogen storage [82-84]. Treatment of myoblasts for 24 h with IGF1 stimulated Prkab.2.82.Luc activity and further enhanced the activation by both ERR isoforms (Fig. 6H). Collectively, these results validate consensome predictions that genomic targets encoding AMPK and PP1 regulatory subunits are under direct transcriptional regulation by ERRs, supporting further studies into a physiological role for ERRs in regulation of glycogen metabolism in skeletal muscle.

BioRxiv DRAFT

Discussion

Receptors, enzymes and transcription factors connect metabolic signals to their biological endpoints through a series of interdependent interactions that are commonly referred to as “signaling pathways”. These three categories of pathway node act as points of convergence and integration on the one hand, and divergence and distribution on the other, to ensure an appropriate response of any given cell to its afferent metabolic cues [1]. The vast majority of information on signaling pathways that is readily available to researchers is canonical in nature and derived from the published literature. Although transcriptomic and ChIP-Seq datasets involving manipulation of these nodes have the potential to provide for the generation of focused hypotheses to resolve mechanistic blind spots in such knowledge, deficits in their management have complicated such re-use [4]. To address this problem, we designed here a knowledgebase, the SPP, which allows bench researchers to routinely evaluate transcriptomic or ChIP-Seq dataset evidence for regulatory relationships between cellular signaling pathway nodes and their downstream targets. To enhance discovery using SPP, we surveyed across these datasets in an unbiased and systematic manner, to generate consensus node-target signatures that would allow researchers to infer and model candidate signaling pathways operating in their biological system of interest. The SPP resource is predicated on the idea that receptors, enzymes, transcription factors and other regulatory nodes are molecular free agents whose function is not necessarily tied to any single context and, by extension, have the theoretical potential to associate in any modular combination in a given cellular context.

The direction and magnitude of regulation of a genomic target by a signaling node are highly contextual considerations that change dynamically in response to a broad spectrum of variables. The principle of the transcriptomic consensomic approach is predicated upon initially suppressing such parameters in favor of a ranking that emphasizes the relative responsiveness of a given gene to regulation by a specific signaling node in a given biosample context. Once a

BioRxiv DRAFT

specific consensome has been retrieved, the user can then evaluate evidence across all the underlying data points to develop a well-informed hypothesis that can be mechanistically validated or refined at the bench. We hypothesized that the CPV correlated with the strength of the mechanistic connection between a node and a given target in a given organ context. Put another way, the more exquisitely interdependent a node-target relationship, the more frequently perturbation of the former would impact the latter. We intend the term “consensome” to be interpreted as a general term to refer to establishing consensus across a set of experiments related by perturbation of a given cellular signaling node. The actual method used to establish such consensus is determined by a variety of factors, such as the type of ‘omics platform, or more practical considerations such as the format of the available data. Indeed, within this study itself we employed different approaches to generating the transcriptomic and ChIP-Seq consensomes that were influenced by the format of the available data.

The SPP resource is characterized by a unique combination of features. The Regulation Report organizes query results according to community classifications of the major categories of signaling pathway module, allowing researchers to readily place the results in context. In addition, previous transcriptomic meta-analysis approaches in the field of cellular signaling are perturbation-centric, and applied to experiments involving a single unique perturbant [85, 86]. Consensomic analysis differs from these approaches in that it is node-centric: that is, it is predicated upon the functional relatedness of *any* genetic or small molecule manipulation of a given pathway node, and allows experiments to be grouped for meta-analysis accordingly. In doing so, it lends the meta-analysis greater statistical power, and calls potential node-target relationships with a higher degree of confidence than would otherwise be possible. A third unique aspect is that other primary analysis and meta-analysis studies describing integration of transcriptomic and ChIP-Seq datasets, although insightful, are limited in scope and exist only as stand-alone literature studies. Ours is the first meta-analysis to be sustainably integrated into an

BioRxiv DRAFT

actively-biocurated public web resource in a manner supporting routine use by researchers lacking formal informatics training. Moreover, the continuous incorporation of newly biocurated datasets and reversioning of consensomes over time will have the effect of iteratively suppressing inter-dataset noise and enhancing the resolution of the true biological signal of a given pathway node in a given biosample. Finally, SPP is to our knowledge the first public resource to provide for gene level integration of transcriptomic and ChIP-Seq datasets mapped to common signaling pathway nodes. Given the highly contextual and nuanced nature of transcription factor-promoter relationships therefore, we anticipate that the ability to dissect experimental factors underlying the consensomes in side-by-side transcriptomic and cistromic/ChIP-Seq Regulation Reports will be of considerable value to users in modeling the specific regulatory mechanisms underlying node-target relationships.

Our resource has a number of limitations, some of these are generic in nature and not specific to SPP. For example, SPP is based upon transcriptomic and ChIP-Seq data since these are by far the most informatically mature and numerous of the various types of archived ‘omics data. Although archived proteomic and metabolomic datasets would improve the resource, those that involve manipulation of mammalian cellular signaling nodes not yet reached a volume where their integration would repay the biocurational effort involved in their integration. Moreover, in an ideal world, there would exist an even distribution among archived ‘omics datasets of node manipulations in organ biosamples. Financial realities dictate however that research is directed towards nodes that show the greatest apparent promise for improving human health, resulting in the skewing of research funding towards those molecules, and leaving other families of signaling nodes either poorly characterized or entirely unstudied. Other limitations of the consensomes relate to the design of available archived experiments. For example, certain targets may be regulated by a given node only under specific circumstances (e.g. acute BSM administration) and if such experiments do not exist or are unavailable, these targets would not

BioRxiv DRAFT

rank highly in the corresponding node consensome. Moreover, a low ranking for a target in a consensome does not necessarily imply the complete absence of a regulatory relationship, and may reflect the requirement for a quite specific cellular context for such regulation to take place. Finally, since SPP is based upon transcriptional methodologies, effects exerted by signaling pathway nodes at the protein level, such as enhanced stabilization or degradation of protein, or modulation of the rate of translation, will not be reflected at the mRNA level.

To maximize statistical power, consensomes in the initial version of SPP encompass datasets involving genetic and pharmacological manipulation of nodes within a given family, and many are possible only by incorporating datasets in biosamples representing numerous distinct organs. Future rates of dataset generation and archiving permitting, node- and organ-specific consensomes of reasonable statistical power will become possible, allowing for more detailed dissection of node- and tissue-specific patterns of transcriptional regulation. Validation of the consensomes relied heavily on use of evidence in the literature. In an ideal future, 'omics datasets and the literature would exist in a mutually enhancing relationship, the former providing researchers with insights that are limited in resolution but broad in scope, the latter providing the focused mechanistic and functional detail required to properly interpret and contextualize the node-target relationships. Paramount to such a scenario is equal ease of access to both the literature and 'omics datasets, such that hypotheses can be generated from 'omics datasets as readily and intuitively as abstracts can be accessed through literature search engines. Moreover, in an era of tightening research budgets, there is a pressing responsibility on the biomedical research community to re-purpose existing assets to allow bench researchers to routinely generate future research hypotheses. An important next step therefore will be to establish interoperability between SPP and knowledgebases such as Reactome [2] that are based upon expert manual curation of the research literature. The high degree of orthogonality

BioRxiv DRAFT

between such initiatives will afford users a more complete perspective on cellular signaling pathways than is currently possible.

BioRxiv DRAFT

Acknowledgements

This work was supported by the National Institute of Diabetes, Digestive and Kidney Diseases (DK097748, DK48807, DK107535, DK56338, DK095686 and DK105126), the National Institute of Child Health and Development (DK097748), the National Cancer Institute (CA125123, CA224260) and the National Heart, Lung and Blood Institute (HL127624). Additional funding was provided by the Dan L. Duncan NCI Comprehensive Cancer Center at Baylor College of Medicine. Microscopic analysis was supported by the Integrated Microscopy Core at Baylor College of Medicine with funding from the Cancer Prevention CPRIT (RP150578) and the John S. Dunn Gulf Coast Consortium for Chemical Genomics. We thank Dr Bethany Hazen and Dr Erin Brown for technical help. Finally we extend our thanks to all investigators who archived their ‘omics datasets, without whose due diligence the Signaling Pathways Project would not be possible.

BioRxiv DRAFT

Figure Titles and Legends

Fig. 1. Major signaling pathway module category classifications in the Signaling Pathway Project knowledgebase. Stable community-endorsed classifications for cellular receptors (International Union of Pharmacology, IUPHAR), enzymes (International Union of Biochemistry and Molecular Biology, IUBMB) and transcription factors (TFClass [19]) make up the foundation of the Signaling Pathways Project data model. 5OHTRs, 5 hydroxytryptamine receptors; LDL, low density lipoprotein; NRs, nuclear receptors. For purposes of clarity, omitted from the transcription factors sunburst are factors with > 3 adjacent zinc fingers (482 genes), Hox-related factors (180 genes), multiple dispersed zinc finger factors (140 genes) and other factors with up to three adjacent zinc fingers (24 genes).

Fig. 2. Key elements of the SPP query and reporting interface. A. Ominer query form. B. The transcriptomic Regulation Report. The default display for single gene queries is by Category, which can be adjusted to cluster data points by biosample or species. The default display for multi-gene queries is by Target. C. The cistromic Regulation Report. IP antigens are identified using case-sensitive AGSs to denote experiments in different species. D. Fold Change information windows for transcriptomic (upper) and cistromic (lower) Regulation Reports display essential information on the data point. E. The Bioactive Small Molecule window displays the pharmacology of any BSMs used in the experiment. F. The Fold Change Detail window places the data point in the context of the wider experiment and dataset, and provides for citation of the dataset.

Fig. 3. Consensome user interface. The example shows genomic targets most frequently significantly differentially expressed in response to genetic or pharmacological manipulation of the human insulin receptor in a transcriptomic experiment. Targets are ranked by default by the consensome P value (CPV), which equates to the probability that the observed frequency of

BioRxiv DRAFT

differential expression occurs by chance. Target symbols link to a SPP Regulation Report filtered by the consensome category and biosample parameters to show the underlying data points.

Fig. 4. Consensome intersection analysis: targets encoding key hepatic metabolic enzymes are redundantly transcriptionally regulated by multiple signaling nodes. A. Schematic representing consensome intersection analysis of the 90th percentiles of the hepatic transcriptomic consensomes for All nodes (All nodes-Mm-liver-TC90), FXR (FXR-Mm-liver-TC90), GR (GR-Mm-liver-TC90) and PPARs (PPARs-Mm-liver-TC90). B. The intersection of the FXR, GR and PPAR mouse hepatic transcriptomic consensome 90th percentiles is enriched for genes encoding enzymes that regulate critical checkpoints in the metabolism of carbohydrates, lipids & cholesterol, amino acids, proteoglycans, heme and other biomolecules. Bold indicates pathway first, rate limiting or branchpoint steps. Asterices indicate gene products deficient or mutated in a known metabolic disorder. See Table 4 for references. Abbreviations: AcCoA, acetyl-CoA; ApoB, apolipoprotein B; FFAs, free fatty acids; G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; UDP-Glc, uridine diphosphate glucose.

Fig. 5. Validation of ER regulation of *TPD52L1* (A-C) and AR regulation of *MBOAT2* (D, E) A) Q-PCR analysis of dose dependent induction by 17BE2 in MCF-7 cells of targets with elevated rankings in the ER-Hs-MG transcriptomic consensome. Cells were treated for 18 h with varying concentrations of 17BE2 alone or 1 nM 17BE2 in combination with 100 nM FULV. Each number is representative of –log[17BE2] such that the number 9 is equivalent to 1 nM 17BE2. Data are representative of three independent experiments. B) MCF-7 cells were immunolabeled with TPD52L1 antibody (green) and imaged by deconvolution widefield microscopy. Images shown are max intensity projections, where DAPI (blue) stains DNA. Scale bar is 10μm. (in the inset, 5μm). M, membrane; N, nucleus; P, perinuclear junctions; SF, stress fibers. C. Depletion of TPD52L1 restricts MCF-7 cell viability. D. Induction of *MBOAT2* in

BioRxiv DRAFT

LNCaP prostate epithelial cells upon treatment with 0.1 nM R1881. E. AR-stimulated viability of LNCaP cells is enhanced by depletion of *MBOAT2.* Cells were harvested on Day 5. Gene expression of *KLK3* and *FKBP5*, known canonical AR target genes, was slightly reduced or unaffected, respectively, by *MBOAT2* siRNA knockdown (data not shown). Statistical significance was determined using PRISM by One-way ANOVA with Tukey's multiple comparison test. * $p < 1E-04$.

Fig. 6. GR and ERR control glycogen synthase activity via regulation of genes encoding PP1 and AMPK regulatory subunits. A. Schematic depiction of tandem regulation of glycogen synthase phosphorylation status and activity by PP1 and AMPK. **B.** GR/NR3C1 ChIPseq data from inguinal adipose tissue (iWAT) showing two prominent peaks 5' to the first exon of *PPP1R3C* (gene is transcribed right to left in this image) and potential GRE within the first GR/NR3C1 peak. **C.** qPCR of glycogenic genes including *Ppp1r3c*, *Ppp1r3b* and the established GR/NR3C1 targets pyruvate carboxylase (*Pcx*) and *Fgf21*. **D.** ESRRA binds to chromatin in the vicinity of the *PRKAB2* gene. **E.** Activity of the human *PRKAB2* promoter-reporter in C2C12 myoblasts (MB) or day 4 myotubes (MT) cotransfected with vector, ERRα/Esrra, ERRγ/Esrrg, or PGC1α/ Ppargc1a, as indicated. **F.** Activity of the *PRKAB2* promoter in MB cultured in 0.1% FBS overnight -/+ 10nM IGF1 treatment for 24 hours. Data are reported as mean luciferase/renilla values normalized to control (± S.E.M.) for three trials. Asterisks indicate significant differences between transfection conditions (*) or IGF1 treatment (**), ($p < 0.05$, n=3). **G.** Expression of endogenous *Prkab2* transcript was measured by quantitative real-time PCR in day 3 C2C12 myotubes (MT) treated with vehicle, 5 μM XCT790 (IC50~0.5 mM) or 0.1 μM DY40 (IC50~10 nM) for 24 h. *Prkab2* transcript levels were normalized to 36B4 expression and results are expressed as the mean ± S.E.M. Asterices * indicate significant difference vehicle vs. treatment groups, ($p \leq 0.05$, n=3). **H.** Quantitative real-time PCR analysis of basal *Prkab2* expression in vastus lateralis muscles of male wild-type

BioRxiv DRAFT

(WT) or ERRα-/- mice (ERRα-/-). *Prkab2* transcript was normalized to 36B4 expression and results are expressed as the mean (± S.E.M). Asterisk * indicates significant difference between groups ($p < 0.05$, n=4).

BioRxiv DRAFT

Table 1: Examples of Single Gene queries in the Signaling Pathways Project knowledgebase. The Ominer query form accommodates any level of detail required, from broad discovery queries across multiple nodes and organs to specific regulatory contexts at lower differential expression cut-offs.

	Signaling Pathway Module			Biosample		
'Omics Category	**Category**	**Class**	**Family**	**System**	**Organ**	**Target**
Transcriptomic	All			All		*PDK4*
	All			Metabolic	Liver	*ALDH3A2*
	Receptor	Catalytic	Insulin receptor	All		*PFKFB3*
	Enzymes	Kinases	Cyclin-dependent	All		*DHRS1*
ChIP-Seq	All			All		*LHPP*
	All			Male Repro	Prostate	*TMPRSS2*
	Enzymes	Acetyltransferase	CBP/p300	All		*MAMDC2*
	Transcription factors	BZIP	C/EBP family	Immune	Leukocytes	*TRIB1*

Table 2: Examples of GO Term queries in the SPP knowledgebase.

	Signaling Pathway Module			
'Omics Category	**Category**	**Class**	**Family**	**GO Term**
Transcriptomic	Receptors	Catalytic	All	Fatty acid beta oxidation
	Receptors	G-protein coupled	All	Glycolytic process
	Enzymes	Kinases	Cyclin-dependent	Adipose tissue development
	Enzymes	Acetyltransferases	All	Inflammatory response
Cistromic (ChIP-Seq)	Receptors	Nuclear	PPAR	Cellular response to fatty acid
	Enzymes	E3 Ubiquitin ligases	All	Carbohydrate biosynthetic process
	Transcription factors	BHLH	All	Gluconeogenesis
	Transcription factors	BZIP	All	Urea cycle

BIORXIV DRAFT

Table 3: Examples of consensomes in the SPP knowledgebase. Consensomes are calculated either at the signaling pathway node family level, or across all experiments in a given organ biosample, to indicate frequently regulated targets in a given organ.

	Signaling Pathway Module			Biosample		
'Omics	**Category**	**Class**	**Family**	**System**	**Organ**	**Species**
Transcriptomic	Receptors	G protein-coupled	Adrenergic	All		Mouse
		Catalytic	Chemokine	All		Mouse
			Toll-like	All		Mouse
			Insulin receptor	All		Human
			Leptin receptor	All		Mouse
		Nuclear	PPARs	All		Mouse
			Glucocorticoid	All		Mouse
	Enzymes	Kinases	Cyclin-dependent	All		Human
	All			Metabolic	Liver	Mouse
	All			Metabolic	Adipose	Mouse
ChIP-Seq	Enzymes	Acetyltransferases	CBP/p300	All		Human, Mouse
			NCOA family	All		Human, Mouse
	Transcription	BZIP	C/EBP family	All		Human, Mouse
		E2F/FOX	E2F family	All		Human, Mouse
			FOXO family	All		Human, Mouse

BioRxiv DRAFT

Table 4. Rate limiting, committed step and metabolic disease deficiency-associated hepatic metabolic enzymes in the intersection of the 90th percentiles of the All nodes, GR, FXR and PPAR family liver transcriptomic consensomes

Target	Gene product	Hepatic metabolic pathway	Known human deficiency disease
Lipid metabolism			
*Acaa1a**	3-ketoacyl-CoA thiolase A, peroxisomal	Beta-oxidation of fatty acids	ACAA1, Pseudo-Zellweger [87]
Acaca*	Acetyl-CoA carboxylase alpha	Rate limiting in long chain fatty acid synthesis [37]	ACACA [38]
Acly	ATP citrate lyase	Rate limiting in *de novo* lipogenesis [88]	
Acsl5	Acyl-CoA synthetase long chain family member 5	Branch point enzyme in fatty acid metabolism [89]	-
Agpat6	Acylglycerolphosphate acyltransferase 4	Rate limiting in *de novo* triacylglycerol biosynthesis [90]	
*Arsa**	Arylsulfatase A	Sulfatide biosynthesis	Metachromatic leukodystrophy [91]
Fads2*	Fatty acid desaturase 2	Rate limiting in biosynthesis of unsaturated fatty acids	Δ6-fatty acid desaturase [92]
Gpam	Glycerol-3-phosphate acyltransferase, mitochondrial	Initial and committed step in glycerolipid biosynthesis [93]	
*Hadh**	Hydroxyacyl-coenzyme A dehydrogenase	Catalyzes the mitochondrial oxidation of straight-chain 3-hydroxyacyl-CoAs as part of the beta-oxidation pathway	Hyperinsulinemic hypoglycemia [94]
Mgll	Monoglyceride lipase	Rate limiting in monoacylglycerol catabolism [95]	
*St3gal5**	ST3 beta-galactoside alpha-2,3-sialyltransferase 5	Ganglioside biosynthesis	Salt & Pepper Syndrome [96]
Carbohydrate metabolism			
Entpd5	Ectonucleoside triphosphate diphosphohydrolase 5	Promotes glycolysis & Warburg effect [97]	-
Got1	Aspartate aminotransferase, cytoplasmic	Transamination of aspartate & oxaloacetate in gluconeogenesis	-
Parp14	Poly [ADP-ribose] polymerase 14	Promotes Warburg effect [41]	
*Pklr**	pyruvate kinase L/R	Catalyzes the transphosphorylation of phohsphoenolpyruvate into pyruvate and ATP, the rate-limiting step of glycolysis	Chronic hereditary nonspherocytic hemolytic anemia
Ppp1r3c*	Protein phosphatase 1 regulatory subunit 3C	Regulatory subunits of PP1, a direct regulator of glycogen synthase, rate limiting step in glycogen synthesis [98]	Lafora disease [99]
Ppp1r14b*	Protein phosphatase 1 regulatory subunit 14B		

BIORXIV DRAFT

Amino acid metabolism			
*Gpt2**	Glutamate pyruvate transaminase2	Reversible transamination of alanine and α-ketoglutarate to form pyruvate and glutamate	Autosomal receessive mental retardation
Hal*	Histidine ammonia-lyase	Initial reaction in histidine catabolism	Histidemia [39]
Tdo2*	Tryptophan 2,3-dioxygenase	Initial and rate limiting step in the kynurenine tryptophan catabolism pathway [100]	Hypertryptophanemia [101]
Cholesterol metabolism			
Acat2*	Acetyl-CoA acetyltransferase 2	Rate limiting in cholesterol esterification [102]	ACAT2 [103]
Dhcr7*	7-dehydrocholesterol reductase	Terminal enzyme in Kandutsch-Russell cholesterol synthesis [104]	Smith-Lemli-Opitz syndrome [105]
Other metabolic pathways			
Adh4	Alcohol dehydrogenase	Rate limiting step in alcohol metabolism [106]	
*Alad**	Aminolevulinate dehydratase	Porphobilinogen biosynthesis	Acute hepatic porphyria [107]
Aldh1a1*	Aldehyde dehydrogenase 1 family member A1	Irreversible conversion of RALD to ATRA [108]	Gorlin syndrome [109]
*Gamt**	Guanidinoacetate N-methyltransferase	Creatine biosynthesis	Creatine [110]
Hsd3b2*	3beta-hydroxysteroid dehydrogenase/delta(5)-delta(4)isomerase type II	Rate limiting in aldosterone production [111]	
Nnmt	Nicotinamide N-methyltransferase	Methylation of N-adenosyl,ethionine; central role in hepatic glucose and lipid metabolism [40]	

BioRxiv DRAFT

Materials and Methods

Statistical analysis

Full descriptions of the statistical analyses for each experiment are included in the descriptions of those experiments below and in the Figure Legends. A full description of the statistical basis of the consensomes is included in File S1.

Data availability

All SPP datasets and consensomes are freely available on the SPP knowledgebase under a Creative Commons Attribution 3.0 license, which provides for sharing, adaptation and both non-commercial and commercial re-use, as long as the resource is cited.

Signaling Pathways Project web application

The Signaling Pathways Project knowledgebase is a gene-centric Java Enterprise Edition 6, web-based application around which other gene, mRNA, protein and BSM data from external databases are collected. All software is freely available at www.github.com/BCM-DLDCC/nursa. After undergoing semiautomated processed and biocuration as described above, the data and annotations are stored in SPP's Oracle 12c database. RESTful web services expose SPP data, which are served to responsively designed views in the user interface, were created using a Flat UI Toolkit with a combination of JavaScript, D3.JS, AJAX, HTML5, and CSS3. JavaServer Faces and PrimeFaces are the primary technologies behind the user interface. SPP has been optimized for Firefox 24+, Chrome 30+, Safari 5.1.9+, and Internet Explorer 9+, with validations performed in BrowserStack and load testing in LoadUIWeb. XML describing each dataset and experiment is generated and submitted to CrossRef to mint DOIs. Programmatic access through API Application programming interface (API) documentation is available on the Signaling Pathways Project knowledgebase at www.signalingpathways.org/spp/rs/index.jsf .

BioRxiv DRAFT

Node and biosample category mapping

Node mappings were adapted from existing, mature classifications for receptors (International Union of Pharmacology, IUPHAR; [17]), enzymes (International Union of Biochemistry and Molecular Biology Enzyme Committee [18]) and transcription factors (TFClass [19]). To resolve differences between these classifications with respect to the number of hierarchical tiers, and to facilitate the design of the data models, each was reduced to a four-levels Category, Class, Family and Node as shown in Table S1. Biosample category mappings were carried as previously described [15]. To enhance the interoperability of SPP with other databases and pathway resources, small molecule-receptor mappings were based upon those maintained by the International Union of Pharmacology Guide to Pharmacology [17], a pharmacology community biocuration authority.

Consensomes

For cistromic consensomes, MACS2 peak calls from the ChIP-Atlas resource [22] for all nodes in a defined SPP family were averaged and the targets ranked based upon this value. For transcriptomic consensomes, differential expression values and associated significance measures were generated from appropriate experimental contrasts in GEO Series as previously described [112]. Consensomes were generated on a computer cluster and stored in the SPP Oracle 12c database

Maintenance and versioning of consensomes

SPP is continually expanding its base of data points by adding newly biocurated datasets to the resource. Accordingly, a quarterly process identifies all node/family and biosample category combinations represented by datasets added in the previous quarter and calculates new versions of the corresponding consensomes. A statement above the scatterplot and contained

BioRxiv DRAFT

in the associated spreadsheet identifies the specific combination of pathway node, biosample (physiological system and organ) and species represented by the consensome, the version and date stamp, and the total number of data points, experiments and datasets on which it is based.

Bench validation and characterization experiments

Validation and characterization of *TPD52L1* in the ER human mammary gland consensome

ER-Hs-mammary gland transcriptomic consensome Q-PCR MCF-7 cells were maintained in DMEM and Ham/F12 Nutrient Mixture (DMEM/F12) supplemented with 8% Fetal Bovine Serum (FBS), Sodium Pyruvate (NaPyr) and non-essential amino acids (NEAA) and passaged every 2-3 days. For experiments, cells were plated in media lacking phenol red with 8% charcoal-stripped FBS (CFS; Gemini). Cells were plated for 48 hours and then treated for 18 hours with 17BE2 (Sigma), or FULV (Tocris). Total RNA was isolated using the Aurum Total RNA Mini-Kit according to the manufacturer's instructions (Bio-Rad). Total RNA (0.5µg) was reverse-transcribed to cDNA using the iScript cDNA synthesis Kit (Bio-Rad). qPCR was performed using 1.625 µL of Bio-Rad SYBR green supermix, 0.125 µL of a 10 µM dilution of each forward and reverse primer, 0.25 µL of water and 1.25 µL of diluted cDNA for a total reaction volume of 3.25 µL. PCR amplification was carried out using the CFX384 qPCR system. Fold induction was calculated using the $2^{-\Delta\Delta Ct}$ method [113], and normalized to 36B4. All data shown is representative of at least three independent experiments. Primer sequences are shown in Table S4.

Subcellular distribution. MCF-7 cells were kept in 5% CD-CS for 48 hrs prior treatment with 17BE2 10nM for 24 h. A previously published immunofluorescence protocol was followed [114]. Briefly, cells were fixed in 4% formaldehyde in PEM buffer (80 mM potassium PIPES [pH 6.8], 5 mM EGTA, and 2 mM MgCl2), quenched with 0.1 M ammonium chloride for 10 min, and

permeabilized with 0.5% Triton X-100 for 30 min. Cells were incubated at room temperature in 5% Blotto for 1 h, and then specific antibodies were added overnight at 4ºC prior to 30 min of secondary antibody (AlexaFluor488 conjugated; Molecular Probes, 1:1000) and DAPI staining. Primary antibody (rabbit polyclonal, Proteintech 14732-1-AP) was diluted at 1:50. A secondary antibody only control showed no appreciable signal (data not shown). Imaging was performed on a GE Healthcare DVLive image restoration deconvolution microscope using an Olympus PlanApo 40x/0.95NA with z-stacks (0.25µm steps covering 12µm) and deconvolved. Images shown are from a maximum intensity projection.

Cell proliferation assay.

MCF-7 cells (from BCM Tissue Culture Core via ATCC) were plated at $3x10^5$ cells per well of a six well plate in phenol red-free DMEM supplemented with 5% charcoal-stripped FBS. Cells were transfected with 50 nM of a siGENOME SMARTpool targeting human *TPD52L1* (Dharmacon, M-019567-02) or 50 nM of a siGENOME non-targeting pool #2 (Dharmacon, D-001206-14-05) using RNAiMAX (Invitrogen). After two days of knockdown, the cells were split to a 96-well plate in the same media and subsequently treated with (-/+) 10 nM water-soluble 17BE2 (Sigma) for 24 hours. After control or TPD52L1 siRNA transfections and (-/+) 17BE2 treatments, cell viabilities were measured by a CellTiter-Glo® Luminescent assay (Promega). Total RNA was isolated with an RNeasy kit (Qiagen). cDNA was made using 1 µg total RNA and Superscript III reverse transcriptase (Invitrogen) in 20 µl reactions total. To measure the relative mRNA levels, real-time reverse transcription- quantitative PCR (RT-qPCR) was performed in an Applied Biosystems Step One Plus real-time PCR system (Applied Biosystems, Foster City, CA) using 2 µl cDNA diluted 1:10, 900 nM primers, and 0.1 nM Universal Probe designed by the Roche Assay Design Center. Human TPD52L1 primers and probe were forward, 5'-CAACTGTCACAAGCCTCAAGA-3'; reverse, 5'-AGCCTCCTGCCAAGCTCt-3'; Roche probe #73; human β-actin primers and probe were previously described [115]. Average threshold cycle

(Ct) values of human β-actin mRNA were subtracted from corresponding average Ct values of *TPD52L1* mRNA to obtain ΔCt values. Relative mRNA levels were expressed as 2−ΔΔCt compared to the non-targeting siRNA control [116]. Statistical significance was determined using the Student's t-Test, and p values < 0.05 were considered significant.

Validation and characterization of MBOAT2 in the AR Hs prostate consensomes

Cell Culture siRNA Transient Transfections and R1881 Treatments

LNCaP cells (ATCC; Baylor College of Medicine Tissue Culture Core) were plated in 12-well dishes (for gene expression analyses) or 6-well dishes (for cell viability assays) at $1x10^6$ and $2x10^6$, respectively, in charcoal stripped RPMI 1640 media (supplemented with 10% stripped-stripped fetal calf serum, penicillin/streptomycin) and transfected in triplicate with 50 nM of an *MBOAT2* targeting siRNA or a non-targeting siRNA using TransIT-TKO transfection reagent for 5 days. For gene expression analyses, 1 nM R1881 was added to cells on day 4. For cell viability assays, 0.1 nM R1881 was added to cells on day 2. All samples were then harvested on day 5.

Gene Expression Analyses by RT-qPCR

On day 5, the RNA from the 12-well plate LNCaP cell samples was harvested using Tri-reagent, following the manufacturer's instructions. The RNA concentrations were quantitated by Nanodrop (ThermoFisher Nanodrop Lite). 1 ug of each RNA sample was used to make cDNAs by First-Strand cDNA Synthesis using SuperScript II Reverse Transcriptase, following the manufacturer's protocol. cDNAs were then diluted with 180 µl of DEPC-treated water. To analyze gene expression, 2 µl of cDNAs were used in the RT-qPCR reactions along with Taqman Universal MM II, 200 nM primers (using Roche Diagnostics Universal ProbeLibrary System Assay Design *ACTB*: forward 5'-CCAACCGCGAGAAGATGA-3', reverse 5'-CCAGAGGCGTACAGGGATAG-3', probe #64; *MBOAT2*, forward 5'-

BioRxiv DRAFT

TCAGACAGCTCTTTGGCTCA-3', reverse 5'-ACACCCCTGTTAGAAACGTTAGAT-3', probe #53; *KLK3*, forward 5'-CCTGTCCGTGACGTGGAT-3', reverse 5'-CAGGGTTGGGAATGCTTCT-3', probe #75; and *FKBP5*, forward 5'-ACAATGAAGAAAGCCCCACA-3', reverse 5'-CACCATTCCCCACTCTTTTG-3', probe #55,) , on a StepOnePlus machine (Applied Biosystems). Expression levels of *MBOAT2*, *KLK3*, and *FKBP5* were normalized to *ACTB* and determined by the ΔCt method. PRISM software was used for statistical analyses.

Cell Viability Assay

On day 5, the 6-well plate LNCaP cells were briefly trypsinized and collected. Cell viability was then determined using CellTiter-Glo Luminescent Cell Viability Assay, following the manufacturer's instruction, and a Berthold 96 well plate reading luminometer. PRISM software was used for the statistical analyses.

Validation and characterization of *Ppp1r3c* in the GR mouse metabolic consensome

Hepa1c cells were grown in DMEM with 10% fetal bovine serum and penicillin, streptomycin and gentamycin (Life Technologies) and treated with vehicle (ethanol) or 250 nM DEX (Sigma) for 48h. Cells were lysed in TriZOL and total RNA was purified by a PureLink RNA Kit. 250 µg of RNA was reverse transcribed into cDNA using a High Capacity cDNA Reverse Transcription Kit (Life Technologies). Genes were quantified using SYBR Green following the manufacturer's instructions on an QuantStudio 5 qPCR instrument (Applied Biosystems). Gene expression was normalized to an internal control (*Rplp0*; after evaluating several normalization genes to ensure they were unchanged by treatment). Each experiment was standardized to its own vehicle treatment. Primer sequences used are described in Table S4.

Validation and characterization of *Prkab2* in the ERR mouse metabolic consensome

BioRxiv DRAFT

Animals. All animal protocols were approved by the Institutional Animal Care and Use Committee at City of Hope. The ERRα/Esrra-/- mice have been described and were maintained as a hybrid strain (C57BL/6/SvJ129) [117, 118]. For baseline comparisons, littermate wild-type and ERRα/Esrra-/- mice were generated from heterozygous breeders to control for strain background. Skeletal muscle (quadriceps) was isolated from 12 week old mice fed wild-type and ERRα/Esrra-/- mice during the daytime (1000 to 1200 h), flash frozen and stored at -80°C until RNA isolation was performed.

Cell culture and reagents. C2C12 (ATCC, cell line CRL-1772, Manassas, VA) myoblasts (MB) were cultured in growth media (DMEM (Corning Cellgro, Manassas, VA) containing 10% FBS and differentiated in DMEM containing 2% horse serum (Atlanta Biologicals, Lawrenceville, GA) when MB reached confluence. All experiments were performed in cells below passage number 35. C2C12 myocytes were treated with 5μM XCT790 (Sigma-Aldrich, St. Louis, MO), 0.1 μM DY40 [119] or DMSO in growth media or differentiation media prepared with charcoal-stripped serum.

Plasmids and transcriptional activity assays. The Prkab2.-2.82.Luc promoter-reporter contains the region of the mouse *Prkab2* gene encompassing -2815 to +27 bp relative to the predicted TSS. The region was amplified from C57B6/J mouse genomic DNA using primers, 5-CTCGGTACCTGAGCACATTAAACCAGTAGTCC-3; 5-GAGAAGCTTTACAAGGCCCGCGACGAGGTAC-3' (KpnI and HindIII sites in the forward and reverse primers, respectively, denoted in italics) and cloned directly into KpnI/HindIII sites of the pGL3-Basic vector. The entire cloned region was sequenced and confirmed against the corresponding region of the reference Prkab2 gene sequence in NCBI (release 106). The pcDNA3.1-Flag-ERR□, pSG5-HA-ERRα and pcDNA-myc/his-PGC-1α have been previously described [120]. Transient transfection in C2C12 myocytes using the calcium phosphate method and the plasmid concentrations used have been described [79]. Luciferase activity was

assayed in MB 48h post-transfection or in day 4 MT after changing confluent cells to 2% HS/DMEM. To assess IGF1 activation, MB were changed to SFM -/+ 10nM recombinant IGF1 one day following transfection and activities were measured after 24 h treatment. Luciferase activity was assayed using Dual-Glo reagents (Promega, Madison, WI) on a Tecan M200 plate reader (Männedorf, Switzerland). Firefly luciferase activity was normalized to that of renilla luciferase, which was expressed downstream of the minimal thymidine kinase promoter from the pRL-TK-Renilla plasmid.

Quantitative real-time PCR. Real-time PCR was performed to quantify relative transcript levels in RNA collected from skeletal muscle isolated from mice or from day 3 MT using TRIzol reagent (Life Technologies, Carlsbad, CA), as described [79]. RNA (1 µg) was reverse transcribed in 20 µl reactions using the BioRad iScript cDNA Synthesis Kit (BioRad Laboratories) with 1:1 mixture of oligo-dT and random hexamers for 30 min at 42°C. Resulting cDNA is used in PCR reactions (15 µl) performed in 96-well format in triplicate contained 1X SYBR green reagent (BioRad iQ SYBR Green Supermix), 0.4 µM gene specific primers and 0.5 µl of first strand reaction product (diluted 1:2) as previously described [79]. Cycling and detection was performed using BioRad IQ5 Real Time PCR system. Experimental transcript levels were normalized to 36B4 (Rplp0) ribosomal RNA analyzed in separate reactions. The following mouse-specific primer sets were used to detect specific gene expression: AMPKα2 (*Prkab2*) forward, 5- ACCATCTCTATGCACTGTCCA -3; reverse, 5-CAGCGTGGTGACATACTTCTT-3; 36B4 (*Rplp0*) forward, 5-ATCCCTGACGCACCGCCGTGA -3; reverse, 5-TGCATCTGCTTGGAGCCCACGTT-3.

Statistical analysis. All cell experiments were performed in three independent trials with 3 replicates per trial. Data are presented as mean (± S.E.M.) relative activity or expression normalized to control (empty vector or vehicle treated condition). Differences between mean values for luciferase activities and real-time PCR analysis were analyzed by a one-way ANOVA

BioRxiv DRAFT

followed by Fisher's LSD post test or by unpaired Student's t test using PRISM software (GraphPad Software, San Diego, CA). A p-value of ≤0.05 was considered significantly different.

BioRxiv DRAFT

REFERENCES

1. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nature reviews Molecular cell biology. 2006;7(2):85-96. doi: 10.1038/nrm1837. PubMed PMID: 16493415.
2. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic acids research. 2016;44(D1):D481-7. Epub 2015/12/15. doi: 10.1093/nar/gkv1351. PubMed PMID: 26656494; PubMed Central PMCID: PMCPmc4702931.
3. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3:160018. Epub 2016/03/16. doi: 10.1038/sdata.2016.18. PubMed PMID: 26978244; PubMed Central PMCID: PMCPMC4792175.
4. Darlington YF, Naumov A, McOwiti A, Kankanamge WH, Becnel LB, McKenna NJ. Improving the discoverability, accessibility, and citability of omics datasets: a case report. Journal of the American Medical Informatics Association : JAMIA. 2017;24(2):388-93. doi: 10.1093/jamia/ocw096. PubMed PMID: 27413121.
5. McArdle S, Buscher K, Ehinger E, Pramod AB, Riley N, Ley K. PRESTO, a new tool for integrating large-scale -omics data and discovering disease-specific signatures. BioRXiv. 2018. doi: 10.1101/302604 %J bioRxiv.
6. Parikh JR, Klinger B, Xia Y, Marto JA, Bluthgen N. Discovering causal signaling pathways through gene-expression patterns. Nucleic acids research. 2010;38(Web Server issue):W109-17. Epub 2010/05/25. doi: 10.1093/nar/gkq424. PubMed PMID: 20494976; PubMed Central PMCID: PMCPMC2896193.
7. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437-52.e17. Epub 2017/12/02. doi: 10.1016/j.cell.2017.10.049. PubMed PMID: 29195078; PubMed Central PMCID: PMCPMC5990023.
8. Cheneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic acids research. 2018;46(D1):D267-d75. Epub 2017/11/11. doi: 10.1093/nar/gkx1092. PubMed PMID: 29126285; PubMed Central PMCID: PMCPMC5753247.
9. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. Nature communications. 2016;7:12846. Epub 2016/09/27. doi: 10.1038/ncomms12846. PubMed PMID: 27667448; PubMed Central PMCID: PMCPMC5052684.
10. Cantini L, Calzone L, Martignetti L, Rydenfelt M, Bluthgen N, Barillot E, et al. Classification of gene signatures for their information value and functional redundancy. NPJ systems biology and applications. 2018;4:2. Epub 2017/12/22. doi: 10.1038/s41540-017-0038-8. PubMed PMID: 29263798; PubMed Central PMCID: PMCPMC5736638.
11. Schubert M, Klinger B, Klunemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nature communications. 2018;9(1):20. Epub 2018/01/04. doi: 10.1038/s41467-017-02391-6. PubMed PMID: 29295995; PubMed Central PMCID: PMCPMC5750219.
12. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, et al. The Comparative Toxicogenomics Database: update 2017. Nucleic acids research. 2017;45(D1):D972-d8. Epub 2016/09/22. doi: 10.1093/nar/gkw838. PubMed PMID: 27651457; PubMed Central PMCID: PMCPMC5210612.
13. Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, et al. DSigDB: drug signatures database for gene set analysis. Bioinformatics (Oxford, England). 2015;31(18):3069-71. Epub 2015/05/21.

BIORXIV DRAFT

doi: 10.1093/bioinformatics/btv313. PubMed PMID: 25990557; PubMed Central PMCID: PMCPMC4668778.
14. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell systems. 2015;1(6):417-25. Epub 2016/01/16. doi: 10.1016/j.cels.2015.12.004. PubMed PMID: 26771021; PubMed Central PMCID: PMCPMC4707969.
15. Becnel LB, Ochsner SA, Darlington YF, McOwiti A, Kankanamge WH, Dehart M, et al. Discovering relationships between nuclear receptor signaling pathways, genes, and tissues in Transcriptomine. Science signaling. 2017;10(476). doi: 10.1126/scisignal.aah6275. PubMed PMID: 28442630.
16. Thomas SM, Brugge JS. Cellular functions regulated by Src family kinases. Annual review of cell and developmental biology. 1997;13:513-609. doi: 10.1146/annurev.cellbio.13.1.513. PubMed PMID: 9442882.
17. Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SP, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucleic acids research. 2016;44(D1):D1054-68. doi: 10.1093/nar/gkv1037. PubMed PMID: 26464438; PubMed Central PMCID: PMC4702778.
18. Webb EC. Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. . San Diego: Academic Press; 1992.
19. Wingender E, Schoeps T, Donitz J. TFClass: an expandable hierarchical classification of human transcription factors. Nucleic acids research. 2013;41(Database issue):D165-70. doi: 10.1093/nar/gks1123. PubMed PMID: 23180794; PubMed Central PMCID: PMC3531165.
20. Minard AY, Tan SX, Yang P, Fazakerley DJ, Domanova W, Parker BL, et al. mTORC1 Is a Major Regulatory Node in the FGF21 Signaling Network in Adipocytes. Cell reports. 2016;17(1):29-36. doi: 10.1016/j.celrep.2016.08.086. PubMed PMID: 27681418.
21. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. 2016;7:12846. doi: 10.1038/ncomms12846. PubMed PMID: 27667448.
22. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions. bioRxiv. 2018.
23. Foulkes NS, Sassone-Corsi P. More is better: activators and repressors from the same gene. Cell. 1992;68(3):411-4. Epub 1992/02/07. PubMed PMID: 1739963.
24. Bertoli C, Skotheim JM, de Bruin RA. Control of cell cycle transcription during G1 and S phases. Nature reviews Molecular cell biology. 2013;14(8):518-28. Epub 2013/07/24. doi: 10.1038/nrm3629. PubMed PMID: 23877564; PubMed Central PMCID: PMCPmc4569015.
25. Okabe Y, Medzhitov R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. Cell. 2014;157(4):832-44. Epub 2014/05/06. doi: 10.1016/j.cell.2014.04.016. PubMed PMID: 24792964; PubMed Central PMCID: PMCPMC4137874.
26. MacNeill SA. Structure and function of the GINS complex, a key component of the eukaryotic replisome. The Biochemical journal. 2010;425(3):489-500. Epub 2010/01/15. doi: 10.1042/bj20091531. PubMed PMID: 20070258.
27. Hirano T. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. Cell. 2016;164(5):847-57. Epub 2016/02/27. doi: 10.1016/j.cell.2016.01.033. PubMed PMID: 26919425.
28. Chong JP, Mahbubani HM, Khoo CY, Blow JJ. Purification of an MCM-containing complex as a component of the DNA replication licensing system. Nature. 1995;375(6530):418-21. Epub 1995/06/01. doi: 10.1038/375418a0. PubMed PMID: 7760937.

BIORXIV DRAFT

29. Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC. The centromere: hub of chromosomal activities. Science. 1995;270(5242):1591-4. Epub 1995/12/08. PubMed PMID: 7502067.
30. Tsai MJ, O'Malley BW. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. Annual review of biochemistry. 1994;63:451-86. doi: 10.1146/annurev.bi.63.070194.002315. PubMed PMID: 7979245.
31. Wellen KE, Thompson CB. A two-way street: reciprocal regulation of metabolism and signalling. Nature reviews Molecular cell biology. 2012;13(4):270-6. Epub 2012/03/08. doi: 10.1038/nrm3305. PubMed PMID: 22395772.
32. Teodoro JS, Rolo AP, Palmeira CM. Hepatic FXR: key regulator of whole-body energy metabolism. Trends in endocrinology and metabolism: TEM. 2011;22(11):458-66. doi: 10.1016/j.tem.2011.07.002. PubMed PMID: 21862343.
33. Bollen M, Keppens S, Stalmans W. Specific features of glycogen metabolism in the liver. The Biochemical journal. 1998;336 (Pt 1):19-31. PubMed PMID: 9806880; PubMed Central PMCID: PMC1219837.
34. Desvergne B, Michalik L, Wahli W. Transcriptional regulation of metabolism. Physiological reviews. 2006;86(2):465-514. doi: 10.1152/physrev.00025.2005. PubMed PMID: 16601267.
35. Pawlak M, Lefebvre P, Staels B. Molecular mechanism of PPARalpha action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. Journal of hepatology. 2015;62(3):720-33. doi: 10.1016/j.jhep.2014.10.039. PubMed PMID: 25450203.
36. Corcoran CC, Grady CR, Pisitkun T, Parulekar J, Knepper MA. From 20th century metabolic wall charts to 21st century systems biology: database of mammalian metabolic enzymes. American journal of physiology Renal physiology. 2017;312(3):F533-f42. Epub 2016/12/16. doi: 10.1152/ajprenal.00601.2016. PubMed PMID: 27974320; PubMed Central PMCID: PMCPmc5374312.
37. Lopez-Casillas F, Bai DH, Luo XC, Kong IS, Hermodson MA, Kim KH. Structure of the coding sequence and primary amino acid sequence of acetyl-coenzyme A carboxylase. Proceedings of the National Academy of Sciences of the United States of America. 1988;85(16):5784-8. Epub 1988/08/01. PubMed PMID: 2901088; PubMed Central PMCID: PMCPmc281849.
38. Blom W, de Muinck Keizer SM, Scholte HR. Acetyl-CoA carboxylase deficiency: an inborn error of de novo fatty acid synthesis. The New England journal of medicine. 1981;305(8):465-6. doi: 10.1056/NEJM198108203050820. PubMed PMID: 6114432.
39. Virmani K, Widhalm K. Histidinemia: a biochemical variant or a disease? Journal of the American College of Nutrition. 1993;12(2):115-24. PubMed PMID: 8463510.
40. Hong S, Moreno-Navarrete JM, Wei X, Kikukawa Y, Tzameli I, Prasad D, et al. Nicotinamide N-methyltransferase regulates hepatic nutrient metabolism through Sirt1 protein stabilization. Nature medicine. 2015;21(8):887-94. doi: 10.1038/nm.3882

http://www.nature.com/nm/journal/v21/n8/abs/nm.3882.html#supplementary-information.
41. Iansante V, Choy PM, Fung SW, Liu Y, Chai JG, Dyson J, et al. PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation. 2015;6:7882. doi: 10.1038/ncomms8882. PubMed PMID: 26258887.
42. Schmidt-Arras D, Rose-John S. IL-6 pathway in the liver: From physiopathology to therapy. Journal of hepatology. 2016;64(6):1403-15. Epub 2016/02/13. doi: 10.1016/j.jhep.2016.02.004. PubMed PMID: 26867490.
43. Diehl AM. Roles of CCAAT/enhancer-binding proteins in regulation of liver regenerative growth. The Journal of biological chemistry. 1998;273(47):30843-6. Epub 1998/11/13. PubMed PMID: 9812973.

BioRxiv DRAFT

44. Zhao GN, Jiang DS, Li H. Interferon regulatory factors: at the crossroads of immunity, metabolism, and disease. Biochimica et biophysica acta. 2015;1852(2):365-78. Epub 2014/05/09. doi: 10.1016/j.bbadis.2014.04.030. PubMed PMID: 24807060.
45. Wijayaratne AL, McDonnell DP. The human estrogen receptor-alpha is a ubiquitinated protein whose stability is affected differentially by agonists, antagonists, and selective estrogen receptor modulators. The Journal of biological chemistry. 2001;276(38):35684-92. doi: 10.1074/jbc.M101097200. PubMed PMID: 11473106.
46. Shehata M, Weidenhofer J, Thamotharampillai K, Hardy JR, Byrne JA. Tumor protein D52 overexpression and gene amplification in cancers from a mosaic of microarrays. Critical reviews in oncogenesis. 2008;14(1):33-55. PubMed PMID: 19105569.
47. Boutros R, Byrne JA. D53 (TPD52L1) is a cell cycle-regulated protein maximally expressed at the G2-M transition in breast cancer cells. Experimental cell research. 2005;310(1):152-65. doi: 10.1016/j.yexcr.2005.07.009. PubMed PMID: 16112108.
48. JavanMoghadam S, Weihua Z, Hunt KK, Keyomarsi K. Estrogen receptor alpha is cell cycle-regulated and regulates the cell cycle in a ligand-dependent fashion. Cell cycle. 2016;15(12):1579-90. doi: 10.1080/15384101.2016.1166327. PubMed PMID: 27049344; PubMed Central PMCID: PMC4934046.
49. Toyoshima F, Nishida E. Integrin-mediated adhesion orients the spindle parallel to the substratum in an EB1- and myosin X-dependent manner. The EMBO journal. 2007;26(6):1487-98. Epub 2007/02/24. doi: 10.1038/sj.emboj.7601599. PubMed PMID: 17318179; PubMed Central PMCID: PMCPmc1829369.
50. Shehata M, Bieche I, Boutros R, Weidenhofer J, Fanayan S, Spalding L, et al. Nonredundant functions for tumor protein D52-like proteins support specific targeting of TPD52. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(16):5050-60. doi: 10.1158/1078-0432.CCR-07-4994. PubMed PMID: 18698023.
51. Lewis JD, Payton LA, Whitford JG, Byrne JA, Smith DI, Yang L, et al. Induction of tumorigenesis and metastasis by the murine orthologue of tumor protein D52. Molecular cancer research : MCR. 2007;5(2):133-44. doi: 10.1158/1541-7786.MCR-06-0245. PubMed PMID: 17314271.
52. Hayne C, Tzivion G, Luo Z. Raf-1/MEK/MAPK pathway is necessary for the G2/M transition induced by nocodazole. The Journal of biological chemistry. 2000;275(41):31876-82. doi: 10.1074/jbc.M002766200. PubMed PMID: 10884385.
53. Abraham RT. Cell cycle checkpoint signaling through the ATM and ATR kinases. Genes & development. 2001;15(17):2177-96. doi: 10.1101/gad.914401. PubMed PMID: 11544175.
54. Liu Q, Guntuku S, Cui XS, Matsuoka S, Cortez D, Tamai K, et al. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes & development. 2000;14(12):1448-59. PubMed PMID: 10859164; PubMed Central PMCID: PMC316686.
55. Zhang H, Park SH, Pantazides BG, Karpiuk O, Warren MD, Hardy CW, et al. SIRT2 directs the replication stress response through CDK9 deacetylation. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(33):13546-51. doi: 10.1073/pnas.1301463110. PubMed PMID: 23898190; PubMed Central PMCID: PMC3746840.
56. Abbott DW, Holt JT. Mitogen-activated protein kinase kinase 2 activation is essential for progression through the G2/M checkpoint arrest in cells exposed to ionizing radiation. The Journal of biological chemistry. 1999;274(5):2732-42. PubMed PMID: 9915804.
57. Hishikawa D, Shindou H, Kobayashi S, Nakanishi H, Taguchi R, Shimizu T. Discovery of a lysophospholipid acyltransferase family essential for membrane asymmetry and diversity. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(8):2830-5. doi: 10.1073/pnas.0712245105. PubMed PMID: 18287005; PubMed Central PMCID: PMC2268545.

BIORXIV DRAFT

58. Han W, Gao S, Barrett D, Ahmed M, Han D, Macoska JA, et al. Reactivation of androgen receptor-regulated lipid biosynthesis drives the progression of castration-resistant prostate cancer. Oncogene. 2018;37(6):710-21. Epub 2017/10/24. doi: 10.1038/onc.2017.385. PubMed PMID: 29059155; PubMed Central PMCID: PMCPMC5805650.
59. Tsai HC, Boucher DL, Martinez A, Tepper CG, Kung HJ. Modeling truncated AR expression in a natural androgen responsive environment and identification of RHOB as a direct transcriptional target. PloS one. 2012;7(11):e49887. doi: 10.1371/journal.pone.0049887. PubMed PMID: 23209612; PubMed Central PMCID: PMC3510170.
60. Obinata D, Takayama K, Takahashi S, Inoue S. Crosstalk of the Androgen Receptor with Transcriptional Collaborators: Potential Therapeutic Targets for Castration-Resistant Prostate Cancer. Cancers. 2017;9(3). doi: 10.3390/cancers9030022. PubMed PMID: 28264478; PubMed Central PMCID: PMC5366817.
61. Cunha GR, Donjacour AA, Cooke PS, Mee S, Bigsby RM, Higgins SJ, et al. The endocrinology and developmental biology of the prostate. Endocrine reviews. 1987;8(3):338-62. doi: 10.1210/edrv-8-3-338. PubMed PMID: 3308446.
62. He WW, Sciavolino PJ, Wing J, Augustus M, Hudson P, Meissner PS, et al. A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. Genomics. 1997;43(1):69-77. doi: 10.1006/geno.1997.4715. PubMed PMID: 9226374.
63. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, Kim M, Desai N, Young P, et al. Roles for Nkx3.1 in prostate development and cancer. Genes & development. 1999;13(8):966-77. PubMed PMID: 10215624; PubMed Central PMCID: PMC316645.
64. Tabe S, Hikiji H, Ariyoshi W, Hashidate-Yoshida T, Shindou H, Shimizu T, et al. Lysophosphatidylcholine acyltransferase 4 is involved in chondrogenic differentiation of ATDC5 cells. Scientific reports. 2017;7(1):16701. doi: 10.1038/s41598-017-16902-4. PubMed PMID: 29196633; PubMed Central PMCID: PMC5711957.
65. Koelling S, Miosge N. Sex differences of chondrogenic progenitor cells in late stages of osteoarthritis. Arthritis and rheumatism. 2010;62(4):1077-87. doi: 10.1002/art.27311. PubMed PMID: 20131243.
66. Dashty M. A quick look at biochemistry: carbohydrate metabolism. Clinical biochemistry. 2013;46(15):1339-52. Epub 2013/05/18. doi: 10.1016/j.clinbiochem.2013.04.027. PubMed PMID: 23680095.
67. Fong NM, Jensen TC, Shah AS, Parekh NN, Saltiel AR, Brady MJ. Identification of binding sites on protein targeting to glycogen for enzymes of glycogen metabolism. The Journal of biological chemistry. 2000;275(45):35034-9. Epub 2000/08/11. doi: 10.1074/jbc.M005541200. PubMed PMID: 10938087.
68. Jeon SM. Regulation and function of AMPK in physiology and diseases. Experimental & molecular medicine. 2016;48(7):e245. Epub 2016/07/16. doi: 10.1038/emm.2016.81. PubMed PMID: 27416781; PubMed Central PMCID: PMCPmc4973318.
69. Vanstapel F, Dopere F, Stalmans W. The role of glycogen synthase phosphatase in the glucocorticoid-induced deposition of glycogen in foetal rat liver. The Biochemical journal. 1980;192(2):607-12. PubMed PMID: 6263259; PubMed Central PMCID: PMC1162376.
70. Huss JM, Garbacz WG, Xie W. Constitutive activities of estrogen-related receptors: Transcriptional regulation of metabolism by the ERR pathways in health and disease. Biochimica et biophysica acta. 2015;1852(9):1912-27. doi: 10.1016/j.bbadis.2015.06.016. PubMed PMID: 26115970.
71. Printen JA, Brady MJ, Saltiel AR. PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. Science. 1997;275(5305):1475-8. Epub 1997/03/07. PubMed PMID: 9045612.
72. O'Doherty RM, Jensen PB, Anderson P, Jones JG, Berman HK, Kearney D, et al. Activation of direct and indirect pathways of glycogen synthesis by hepatic overexpression of

BioRxiv DRAFT

protein targeting to glycogen. The Journal of clinical investigation. 2000;105(4):479-88. Epub 2000/02/23. doi: 10.1172/jci8673. PubMed PMID: 10683377; PubMed Central PMCID: PMCPmc289167.
73. Dasgupta B, Ju JS, Sasaki Y, Liu X, Jung SR, Higashida K, et al. The AMPK beta2 subunit is required for energy homeostasis during metabolic stress. Molecular and cellular biology. 2012;32(14):2837-48. doi: 10.1128/MCB.05853-11. PubMed PMID: 22586267; PubMed Central PMCID: PMC3416196.
74. Kim J, Yang G, Kim Y, Kim J, Ha J. AMPK activators: mechanisms of action and physiological activities. Experimental & molecular medicine. 2016;48:e224. doi: 10.1038/emm.2016.16. PubMed PMID: 27034026; PubMed Central PMCID: PMC4855276.
75. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, et al. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome research. 2009;19(12):2163-71. doi: 10.1101/gr.097022.109. PubMed PMID: 19801529; PubMed Central PMCID: PMC2792167.
76. Reshef L, Ballard FJ, Hanson RW. The role of the adrenals in the regulation of phosphoenolpyruvate carboxykinase of rat adipose tissue. The Journal of biological chemistry. 1969;244(20):5577-81. PubMed PMID: 5348601.
77. Patel R, Bookout AL, Magomedova L, Owen BM, Consiglio GP, Shimizu M, et al. Glucocorticoids regulate the metabolic hormone FGF21 in a feed-forward loop. Molecular endocrinology. 2015;29(2):213-23. doi: 10.1210/me.2014-1259. PubMed PMID: 25495872; PubMed Central PMCID: PMC4318881.
78. Willy PJ, Murray IR, Qian J, Busch BB, Stevens WC, Jr., Martin R, et al. Regulation of PPARgamma coactivator 1alpha (PGC-1alpha) signaling by an estrogen-related receptor alpha (ERRalpha) ligand. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(24):8912-7. doi: 10.1073/pnas.0401420101. PubMed PMID: 15184675; PubMed Central PMCID: PMC428446.
79. Murray J, Huss JM. Estrogen-related receptor alpha regulates skeletal myocyte differentiation via modulation of the ERK MAP kinase pathway. American journal of physiology Cell physiology. 2011;301(3):C630-45. doi: 10.1152/ajpcell.00033.2011. PubMed PMID: 21562305; PubMed Central PMCID: PMC3174569.
80. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic acids research. 2016;44(D1):D110-5. doi: 10.1093/nar/gkv1176. PubMed PMID: 26531826; PubMed Central PMCID: PMC4702842.
81. Dufour CR, Wilson BJ, Huss JM, Kelly DP, Alaynick WA, Downes M, et al. Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERRalpha and gamma. Cell metabolism. 2007;5(5):345-56. doi: 10.1016/j.cmet.2007.03.007. PubMed PMID: 17488637.
82. Siddle K. Signalling by insulin and IGF receptors: supporting acts and new players. Journal of molecular endocrinology. 2011;47(1):R1-10. Epub 2011/04/19. doi: 10.1530/jme-11-0022. PubMed PMID: 21498522.
83. Ciaraldi TP, Carter L, Rehman N, Mohideen P, Mudaliar S, Henry RR. Insulin and insulin-like growth factor-1 action on human skeletal muscle: preferential effects of insulin-like growth factor-1 in type 2 diabetic subjects. Metabolism: clinical and experimental. 2002;51(9):1171-9. Epub 2002/08/30. PubMed PMID: 12200763.
84. Annunziata M, Granata R, Ghigo E. The IGF system. Acta diabetologica. 2011;48(1):1-9. doi: 10.1007/s00592-010-0227-z. PubMed PMID: 21042815.
85. Jagannathan V, Robinson-Rechavi M. Meta-analysis of estrogen response in MCF-7 distinguishes early target genes involved in signaling and cell proliferation from later target genes involved in cell cycle and DNA repair. BMC systems biology. 2011;5:138. doi: 10.1186/1752-0509-5-138. PubMed PMID: 21878096; PubMed Central PMCID: PMC3225231.

BIORXIV DRAFT

86. Stanislawska-Sachadyn A, Sachadyn P, Limon J. Transcriptomic Effects of Estrogen Starvation and Induction in the MCF7 Cells. The Meta-analysis of Microarray Results. Current pharmaceutical biotechnology. 2015;17(2):161-72. PubMed PMID: 26511976.
87. Goldfischer S, Collins J, Rapin I, Neumann P, Neglia W, Spiro AJ, et al. Pseudo-Zellweger syndrome: deficiencies in several peroxisomal oxidative activities. The Journal of pediatrics. 1986;108(1):25-32. PubMed PMID: 2868085.
88. Zhou Y, Bollu LR, Tozzi F, Ye X, Bhattacharya R, Gao G, et al. ATP citrate lyase mediates resistance of colorectal cancer cells to SN38. Molecular cancer therapeutics. 2013;12(12):2782-91. doi: 10.1158/1535-7163.MCT-13-0098. PubMed PMID: 24132143; PubMed Central PMCID: PMC4302275.
89. Bu SY, Mashek DG. Hepatic long-chain acyl-CoA synthetase 5 mediates fatty acid channeling between anabolic and catabolic pathways. Journal of lipid research. 2010;51(11):3270-80. doi: 10.1194/jlr.M009407. PubMed PMID: 20798351; PubMed Central PMCID: PMC2952567.
90. Wendel AA, Lewin TM, Coleman RA. Glycerol-3-phosphate acyltransferases: rate limiting enzymes of triacylglycerol biosynthesis. Biochimica et biophysica acta. 2009;1791(6):501-6. Epub 2008/11/29. doi: 10.1016/j.bbalip.2008.10.010. PubMed PMID: 19038363; PubMed Central PMCID: PMCPmc2737689.
91. Gieselmann V, Polten A, Kreysing J, von Figura K. Arylsulfatase A pseudodeficiency: loss of a polyadenylylation signal and N-glycosylation site. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(23):9436-40. PubMed PMID: 2574462; PubMed Central PMCID: PMC298511.
92. Stoffel W, Hammels I, Jenke B, Binczek E, Schmidt-Soltau I, Brodesser S, et al. Obesity resistance and deregulation of lipogenesis in Delta6-fatty acid desaturase (FADS2) deficiency. EMBO reports. 2014;15(1):110-20. doi: 10.1002/embr.201338041. PubMed PMID: 24378641; PubMed Central PMCID: PMC4303455.
93. Roy R, Ordovas L, Taourit S, Zaragoza P, Eggen A, Rodellar C. Genomic structure and an alternative transcript of bovine mitochondrial glycerol-3-phosphate acyltransferase gene (GPAM). Cytogenetic and genome research. 2006;112(1-2):82-9. Epub 2005/11/09. doi: 10.1159/000087517. PubMed PMID: 16276094.
94. Molven A, Matre GE, Duran M, Wanders RJ, Rishaug U, Njolstad PR, et al. Familial hyperinsulinemic hypoglycemia caused by a defect in the SCHAD enzyme of mitochondrial fatty acid oxidation. Diabetes. 2004;53(1):221-7. Epub 2003/12/25. PubMed PMID: 14693719.
95. Zechner R, Zimmermann R, Eichmann TO, Kohlwein SD, Haemmerle G, Lass A, et al. FAT SIGNALS--lipases and lipolysis in lipid metabolism and signaling. Cell metabolism. 2012;15(3):279-91. Epub 2012/03/13. doi: 10.1016/j.cmet.2011.12.018. PubMed PMID: 22405066; PubMed Central PMCID: PMCPmc3314979.
96. Boccuto L, Aoki K, Flanagan-Steet H, Chen CF, Fan X, Bartel F, et al. A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. Human molecular genetics. 2014;23(2):418-33. Epub 2013/09/13. doi: 10.1093/hmg/ddt434. PubMed PMID: 24026681; PubMed Central PMCID: PMCPmc3869362.
97. Fang M, Shen Z, Huang S, Zhao L, Chen S, Mak TW, et al. The ER UDPase ENTPD5 promotes protein N-glycosylation, the Warburg effect, and proliferation in the PTEN pathway. Cell. 2010;143(5):711-24. Epub 2010/11/16. doi: 10.1016/j.cell.2010.10.010. PubMed PMID: 21074248.
98. Margolis RN, Cardell RR, Curnow RT. Association of glycogen synthase phosphatase and phosphorylase phosphatase activities with membranes of hepatic smooth endoplasmic reticulum. The Journal of cell biology. 1979;83(2 Pt 1):348-56. PubMed PMID: 227915; PubMed Central PMCID: PMC2111548.

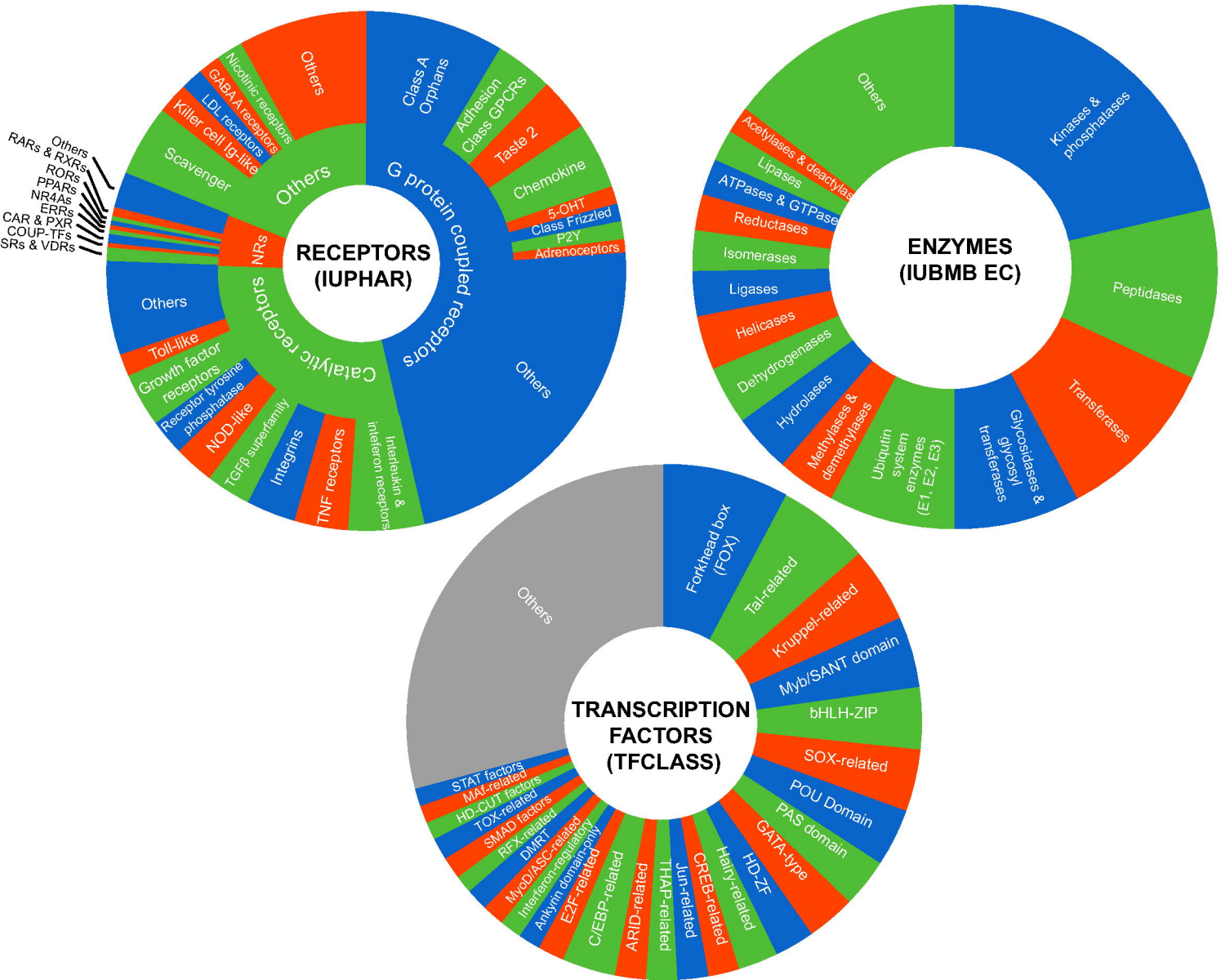
BioRxiv DRAFT

99. Guerrero R, Vernia S, Sanz R, Abreu-Rodriguez I, Almaraz C, Garcia-Hoyos M, et al. A PTG variant contributes to a milder phenotype in Lafora disease. PloS one. 2011;6(6):e21294. doi: 10.1371/journal.pone.0021294. PubMed PMID: 21738631; PubMed Central PMCID: PMC3127956.
100. Efimov I, Basran J, Thackray SJ, Handa S, Mowat CG, Raven EL. Structure and reaction mechanism in the heme dioxygenases. Biochemistry. 2011;50(14):2717-24. Epub 2011/03/03. doi: 10.1021/bi101732n. PubMed PMID: 21361337; PubMed Central PMCID: PMCPmc3092302.
101. Ferreira P, Shin I, Sosova I, Dornevil K, Jain S, Dewey D, et al. Hypertryptophanemia due to tryptophan 2,3-dioxygenase deficiency. Molecular genetics and metabolism. 2017;120(4):317-24. Epub 2017/03/13. doi: 10.1016/j.ymgme.2017.02.009. PubMed PMID: 28285122; PubMed Central PMCID: PMCPmc5421356.
102. Brown JM, Bell TA, 3rd, Alger HM, Sawyer JK, Smith TL, Kelley K, et al. Targeted depletion of hepatic ACAT2-driven cholesterol esterification reveals a non-biliary route for fecal neutral sterol loss. The Journal of biological chemistry. 2008;283(16):10522-34. Epub 2008/02/19. doi: 10.1074/jbc.M707659200. PubMed PMID: 18281279; PubMed Central PMCID: PMCPmc2447638.
103. Bennett MJ, Hosking GP, Smith MF, Gray RG, Middleton B. Biochemical investigations on a patient with a defect in cytosolic acetoacetyl-CoA thiolase, associated with mental retardation. Journal of inherited metabolic disease. 1984;7(3):125-8. PubMed PMID: 6150136.
104. Kandutsch AA, Russell AE. Preputial gland tumor sterols. 3. A metabolic pathway from lanosterol to cholesterol. The Journal of biological chemistry. 1960;235:2256-61. Epub 1960/08/01. PubMed PMID: 14404284.
105. Tint GS, Irons M, Elias ER, Batta AK, Frieden R, Chen TS, et al. Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. The New England journal of medicine. 1994;330(2):107-13. doi: 10.1056/NEJM199401133300205. PubMed PMID: 8259166.
106. Plapp BV. Rate-limiting steps in ethanol metabolism and approaches to changing these rates biochemically. Advances in experimental medicine and biology. 1975;56:77-109. Epub 1975/01/01. PubMed PMID: 167557.
107. Jaffe EK, Stith L. ALAD porphyria is a conformational disease. American journal of human genetics. 2007;80(2):329-37. doi: 10.1086/511444. PubMed PMID: 17236137; PubMed Central PMCID: PMC1785348.
108. Ross AC. Overview of retinoid metabolism. The Journal of nutrition. 1993;123(2 Suppl):346-50. Epub 1993/02/01. PubMed PMID: 8429385.
109. Weber TJ, Magnaldo T, Xiong Y. ALDH1A1 Deficiency in Gorlin Syndrome Suggests a Central Role for Retinoic Acid and ATM Deficits in Radiation Carcinogenesis. Proteomes. 2014;2(3):451-67. doi: 10.3390/proteomes2030451. PubMed PMID: 28250390; PubMed Central PMCID: PMC5302750.
110. Stockler S, Isbrandt D, Hanefeld F, Schmidt B, von Figura K. Guanidinoacetate methyltransferase deficiency: the first inborn error of creatine metabolism in man. American journal of human genetics. 1996;58(5):914-22. PubMed PMID: 8651275; PubMed Central PMCID: PMC1914613.
111. Simard J, Ricketts ML, Gingras S, Soucy P, Feltus FA, Melner MH. Molecular biology of the 3beta-hydroxysteroid dehydrogenase/delta5-delta4 isomerase gene family. Endocrine reviews. 2005;26(4):525-82. doi: 10.1210/er.2002-0050. PubMed PMID: 15632317.
112. Ochsner SA, Watkins CM, McOwiti A, Xu X, Darlington YF, Dehart MD, et al. Transcriptomine, a web resource for nuclear receptor signaling transcriptomes. Physiological genomics. 2012;44(17):853-63. doi: 10.1152/physiolgenomics.00033.2012. PubMed PMID: 22786849; PubMed Central PMCID: PMC3472459.

bioRxiv DRAFT

113. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8. doi: 10.1006/meth.2001.1262. PubMed PMID: 11846609.
114. Stossi F, Dandekar RD, Bolt MJ, Newberg JY, Mancini MG, Kaushik AK, et al. High throughput microscopy identifies bisphenol AP, a bisphenol A analog, as a novel AR down-regulator. Oncotarget. 2016;7(13):16962-74. doi: 10.18632/oncotarget.7655. PubMed PMID: 26918604; PubMed Central PMCID: PMC4941363.
115. Foulds CE, Tsimelzon A, Long W, Le A, Tsai SY, Tsai MJ, et al. Research resource: expression profiling reveals unexpected targets and functions of the human steroid receptor RNA activator (SRA) gene. Molecular endocrinology. 2010;24(5):1090-105. Epub 2010/03/12. doi: 10.1210/me.2009-0427. PubMed PMID: 20219889; PubMed Central PMCID: PMCPmc2870939.
116. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nature protocols. 2008;3(6):1101-8. Epub 2008/06/13. PubMed PMID: 18546601.
117. Huss JM, Imahashi K, Dufour CR, Weinheimer CJ, Courtois M, Kovacs A, et al. The nuclear receptor ERRalpha is required for the bioenergetic and functional adaptation to cardiac pressure overload. Cell metabolism. 2007;6(1):25-37. doi: 10.1016/j.cmet.2007.06.005. PubMed PMID: 17618854.
118. Luo J, Sladek R, Carrier J, Bader JA, Richard D, Giguere V. Reduced fat mass in mice lacking orphan nuclear receptor estrogen-related receptor alpha. Molecular and cellular biology. 2003;23(22):7947-56. PubMed PMID: 14585956; PubMed Central PMCID: PMC262360.
119. Yu DD, Huss JM, Li H, Forman BM. Identification of novel inverse agonists of estrogen-related receptors ERRgamma and ERRbeta. Bioorganic & medicinal chemistry. 2017;25(5):1585-99. doi: 10.1016/j.bmc.2017.01.019. PubMed PMID: 28189393.
120. Huss JM, Kopp RP, Kelly DP. Peroxisome proliferator-activated receptor coactivator-1alpha (PGC-1alpha) coactivates the cardiac-enriched nuclear receptors estrogen-related receptor-alpha and -gamma. Identification of novel leucine-rich interaction motif within PGC-1alpha. The Journal of biological chemistry. 2002;277(43):40265-74. doi: 10.1074/jbc.M206324200. PubMed PMID: 12181319.

BioRxiv DRAFT


RECEPTORS (IUPHAR)
G protein coupled receptors
Class A Orphans
Adhesion Class GPCRs
Taste 2
Chemokine
5-OHT
Class Frizzled
P2Y
Adrenoceptors
Others
Catalytic receptors
Interleukin & inteleron receptors
TNF receptors
Integrins
TGFβ superfamily
NOD-like
Receptor tyrosine phosphatase
Growth factor receptors
Toll-like
Others
NRs
Others
RARs & RXRs
RORs
PPARs
NR4As
ERRs
CAR & PXR
COUP-TFs
SRs & VDRs
Others
Scavenger
Killer cell Ig-like
LDL receptors
GABA A receptors
Nicotinic receptors
Others
ENZYMES (IUBMB EC)
Kinases & phosphatases
Peptidases
Transferases
Glycosidases & glycosyl transferases
Ubiquitin system enzymes (E1, E2, E3)
Methylases & demethylases
Hydrolases
Dehydrogenases
Helicases
Ligases
Isomerases
Reductases
ATPases & GTPases
Lipases
Acetylases & deacetylases
Others
TRANSCRIPTION FACTORS (TFCLASS)
Forkhead box (FOX)
Tal-related
Kruppel-related
Myb/SANT domain
bHLH-ZIP
SOX-related
POU Domain
PAS domain
GATA-type
HD-ZF
Hairy-related
CREB-related
Jun-related
THAP-related
ARID-related
C/EBP-related
E2F-related
Ankyrin domain-only
Interferon-regulatory
MyoD/ASC-related
DMRT
RFX-related
SMAD factors
TOX-related
HD-CUT factors
MAF-related
STAT factors
Others

A.

Start your research

Tip: For best user experience ,click "Reset" to run a new query

Target gene(s) of interest

Single Gene

'Omics Category

Transcriptomics

Start typing and select from the suggested gene symbols

ACOT1 (ACOT1)

Signaling Pathway Module Category

Receptors

Catalytic receptors

Fibroblast growth factor receptors

Biosample Category

All Species

All Physiological Systems

FDR Significance cutoff

5E-02

Reset Submit

B.

Display by: Category

Target
Category
Biosample
Species

Down Up

Transcript Relative Abundance (Fold Change)

-10 0 10

Receptors | Catalytic receptors
Collagen receptor family
SORAF
Epidermal growth factor receptors
AG478
ERLOT
ERBB2
Fibroblast growth factor receptors
SORAF
BFKB1
FGF21
FGF19
Hepatocyte growth factor receptors
PHA665
HGF
IL6 receptor family
IL6
Leptin receptor
LEPR
Toll-like receptors
TLR21L
LPS
Transforming growth factor-β receptors
TGFB1
Tumour necrosis factor receptors
TNF

C.

ChiP-Atlas MACS2 Binding Score(±10 Kb From TSS)

0 500 1000 1500

Enzymes | Deacetylases
Histone deacetylases (HDAC)
Hdac3
HDAC2
HDAC3
Hdac2
Transcription factors | BHLH
Arnt-like factors
Arntl
Clock
Hairy-related
Hes1
Bhlhe40
BHLHE40
HEY1
MESP
TCF21
SREBP factors
Srebf1
SREBF1 | 22RHC, INS
Transcription factors | BZIP
ATF-2-like factors
Atf2
ATF2
ATF-3-like factors
Atf3
ATF3

D.

Fold Change Information

Symbol: Acot1
Fold Change: -29.79
p value: 1.10E-9
Biosample: Metabolic,Liver,whole liver
Experiment: LPS vs Veh (WT)
Species: House Mouse

Bioactive Small Molecule(s)

More Information

FMACS2 Peak

Symbol: ACOT1
MACS2 Binding Score: 1075
MACS2 Q value: 1< E-05
Biosample: Metabolic,Liver,epithelium, HepG2 cells
Experiment: SREBF1 IP | INS + 22RHC - HepG2 cells
Species: Human

Bioactive Small Molecule(s)

More Information

E.

Bioactive Small Molecule(s)

Close

SPP Symbol: 22RHC
BSM Name: 22(R)-Hydroxycholesterol
PubChem CID: 167685
IUPHAR Guide to Pharmacology ID: 2742

SPP Symbol: INS
BSM Name: insulin
PubChem CID:
IUPHAR Guide to Pharmacology ID: 5012

Pharmacology

Signaling Pathway Module Category: Receptors
Class: Catalytic receptors
Family: Insulin receptor family
Node(s): INSR

Signaling Pathway Module Category: Receptors
Class: Nuclear receptors
Family: Liver X receptors
Node(s): NR1H3,NR1H2

Signaling Pathway Module Category: Receptors
Class: Nuclear receptors
Family: Farnesoid X receptor (FXR)
Node(s): NR1H4

F.

Fold Change Details

Close

Fold Change Information

Symbol:	Acot1
Fold Change:	-29.79
p value:	1.10E-9

Experiment Information

Name:	LPS vs Veh (WT)
Description:	Liver was isolated from male BL6/SV129 WT mice treated with 5 µg/gr LPS or vehicle and subsequently fasted for 12 h.
ID:	1
Biosample:	whole liver
Species:	House Mouse

Dataset Information

Name:	Analysis of the acyl-Coenzyme A dehydrogenase, medium chain (Acadm)-dependent and lipopolysaccharide (LPS)-regulated transcriptomes in mouse liver
Description:	Liver was isolated from male BL6/SV129 Acadm KO and WT mice treated with 5 µg/gr LPS or vehicle and subsequently fasted for 12 h.
DOI:	10.1621/kHJ7RMulfF
Download Dataset Citation:	

Consensome (beta)

Download Results

Category: Receptors
Class: Catalytic receptors
Family: Insulin receptor family
Species: Human
Physiological System: All
Organ: All

Consensomes are list of genes ranked according to a meta-analysis of their differential expression in publicly archived transcriptomic datasets involving perturbations of a specific signaling pathway in a given biosample category. Consensome are intended as a guide to identifying those genes most consistently impacted by a given pathway in a given tissue context.

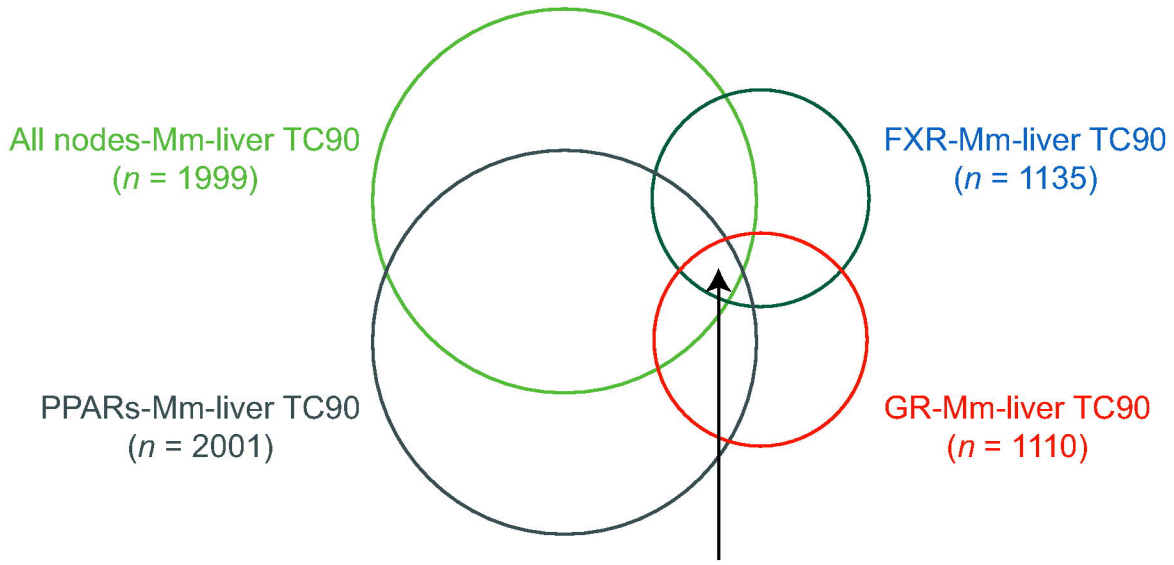
Calculated across **588, 339** data points from **28** experiments in **11** datasets.

Show 50 entries

Search:

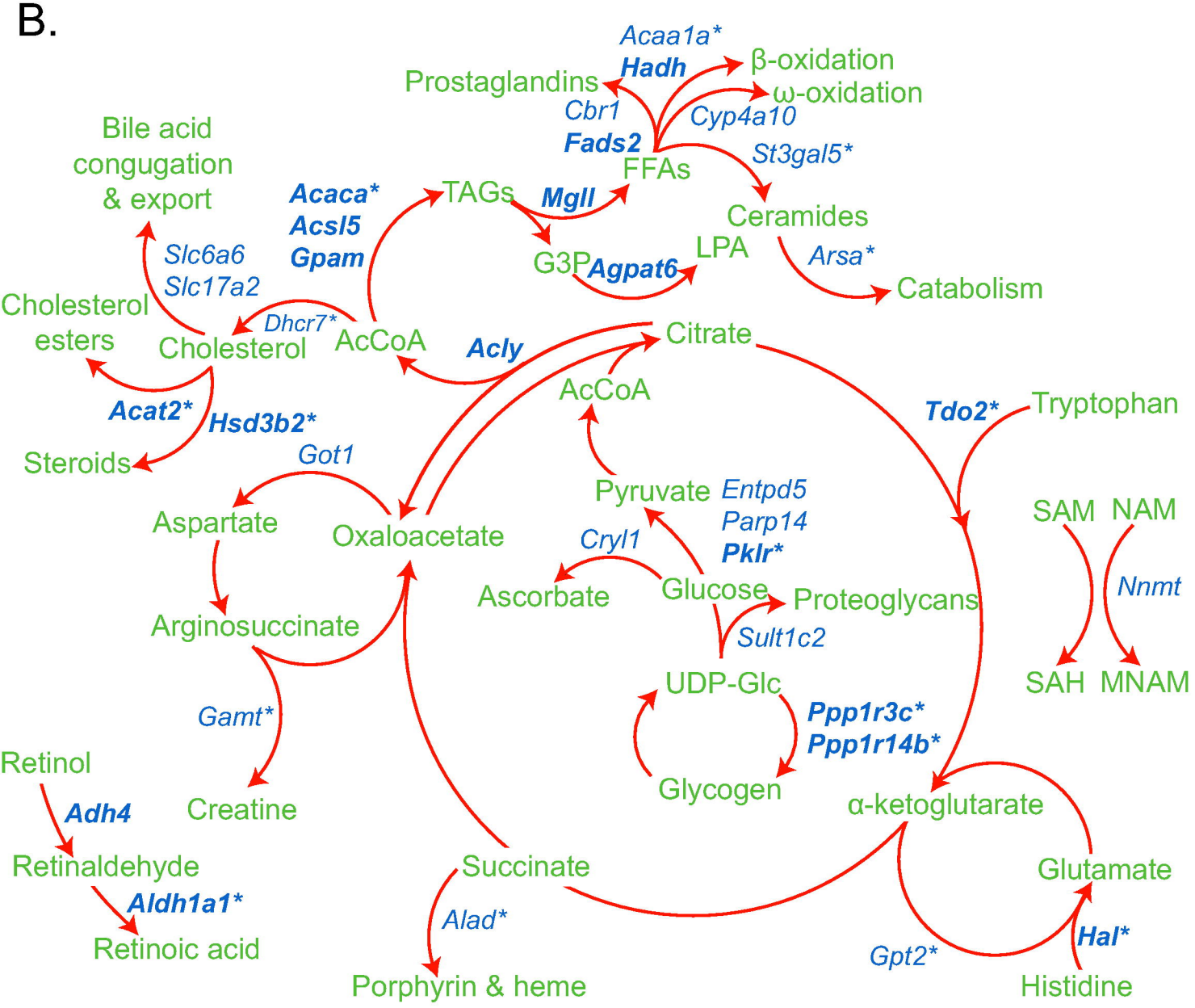
Target	Gene Name	Discovery Rate	GMFC	CPV	Percentile
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	0.913	1.383	1.09E-25	99
GRPEL1	GrpE like 1, mitochondrial	0.913	1.452	1.09E-25	99
GEM	GTP binding protein overexpressed in skeletal muscle	0.87	1.963	1.46E-23	99
CCNG2	cyclin G2	0.87	1.996	1.46E-23	99
ARC	activity regulated cytoskeleton associated protein	0.87	1.619	1.46E-23	99
AVPI1	arginine vasopressin induced 1	0.87	1.384	1.46E-23	99
IER3	immediate early response 3	0.87	1.418	1.46E-23	99
YRDC	yrdC N6-threonylcarbamoyltransferase domain containing	0.87	1.682	1.46E-23	99
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	0.87	2.015	1.46E-23	99
DUSP2	dual specificity phosphatase 2	0.87	1.976	1.46E-23	99
AKIRIN1	akirin 1	0.87	1.51	1.46E-23	99

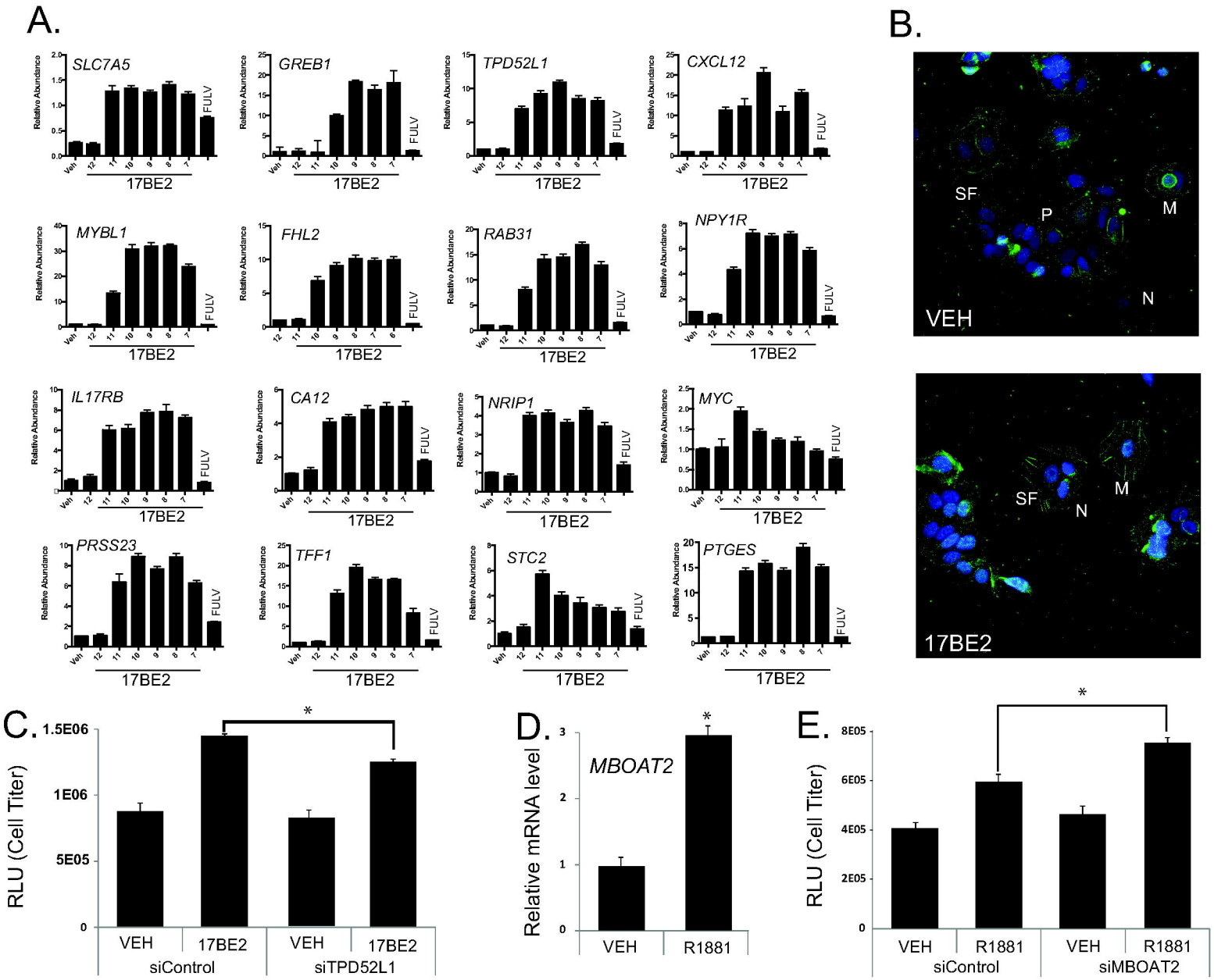
A.



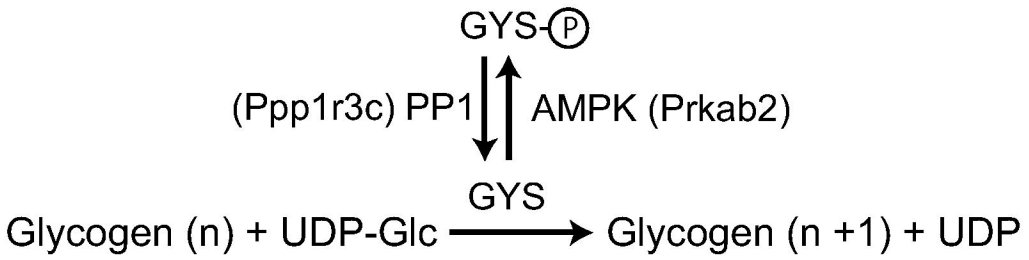
40/117 (34%) intersection targets encode metabolic enzymes

B.

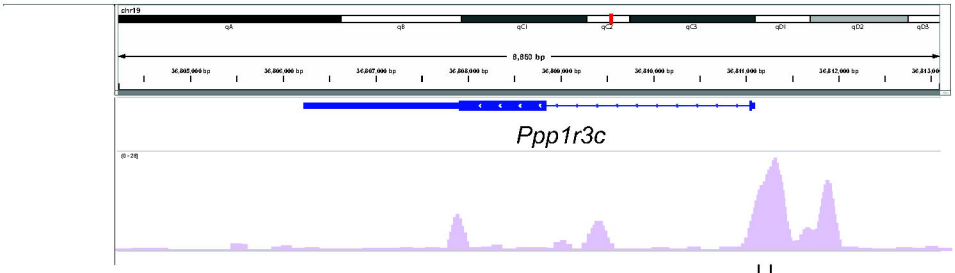



A.
SLC7A5
GREB1
TPD52L1
CXCL12
MYBL1
FHL2
RAB31
NPY1R
IL17RB
CA12
NRIP1
MYC
PRSS23
TFF1
STC2
PTGES
Relative Abundance
Veh
12
11
10
9
8
7
6
FULV
17BE2
B.
SF
P
M
N
VEH
17BE2
C.
RLU (Cell Titer)
1.5E06
1E06
5E05
0
*
VEH
17BE2
VEH
17BE2
siControl
siTPD52L1
D.
MBOAT2
Relative mRNA level
3
2
1
0
*
VEH
R1881
E.
RLU (Cell Titer)
8E05
6E05
4E05
2E05
0
*
VEH
R1881
VEH
R1881
siControl
siMBOAT2

A.

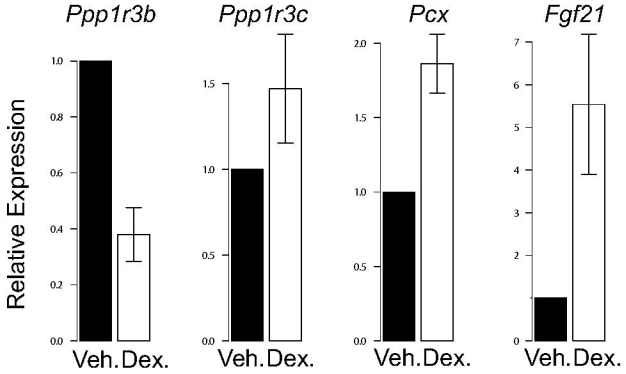


B.

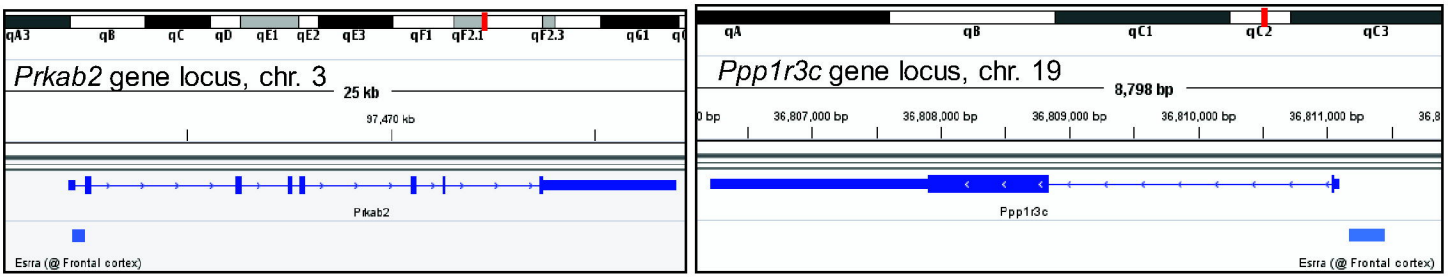


Sequence (5'-3'): GCG**AG**AC**CA**G--**TGT**AGGT
Consensus GRE: XXXAGNACANNNTGTNCTN

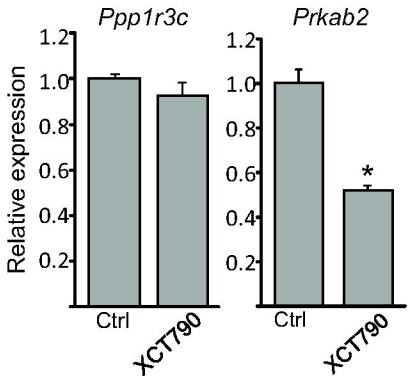
C.



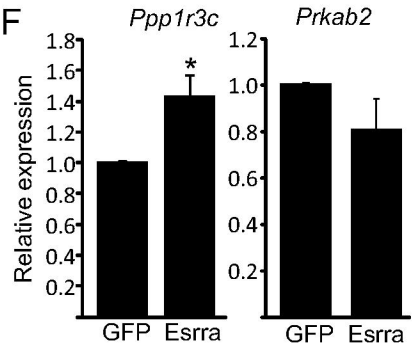
D.



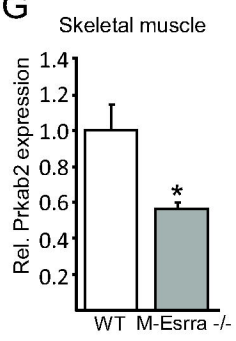
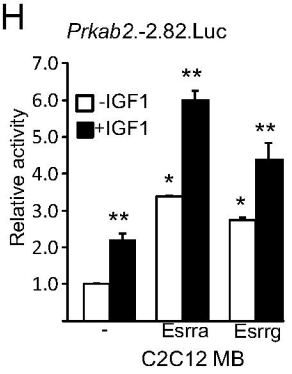
E



F



G



H