




SCIENTIFIC DATA

OPEN
ARTICLE

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

Scott A. Ochsner¹, David Abraham^{1,8}, Kirt Martin^{1,8}, Wei Ding², Apollo McOwiti², Wasula Kankanamge², Zichen Wang ³, Kaitlyn Andreano⁴, Ross A. Hamilton¹, Yue Chen¹, Angelica Hamilton⁵, Marin L. Gantner⁶, Michael Dehart², Shijing Qu², Susan G. Hilsenbeck², Lauren B. Becnel², Dave Bridges⁷, Avi Ma'ayan ³, Janice M. Huss⁵, Fabio Stossi¹, Charles E. Foulds¹, Anastasia Kralli⁶, Donald P. McDonnell⁴ & Neil J. McKenna ^{1*}

Mining of integrated public transcriptomic and ChIP-Seq (cistromic) datasets can illuminate functions of mammalian cellular signaling pathways not yet explored in the research literature. Here, we designed a web knowledgebase, the Signaling Pathways Project (SPP), which incorporates community classifications of signaling pathway nodes (receptors, enzymes, transcription factors and co-nodes) and their cognate bioactive small molecules. We then mapped over 10,000 public transcriptomic or cistromic experiments to their pathway node or biosample of study. To enable prediction of pathway node-gene target transcriptional regulatory relationships through SPP, we generated consensus 'omics signatures, or consensomes, which ranked genes based on measures of their significant differential expression or promoter occupancy across transcriptomic or cistromic experiments mapped to a specific node family. Consensomes were validated using alignment with canonical literature knowledge, gene target-level integration of transcriptomic and cistromic data points, and in bench experiments confirming previously uncharacterized node-gene target regulatory relationships. To expose the SPP knowledgebase to researchers, a web browser interface was designed that accommodates numerous routine data mining strategies. SPP is freely accessible at <https://www.signalingpathways.org>.

Introduction

Signal transduction pathways describe functional interdependencies between distinct classes of molecules that collectively determine the response of a given cell to its afferent endocrine, paracrine and cytokine signals¹. The bulk of readily accessible information on these pathways resides in peer-reviewed research articles and in knowledgebases that curate such information². 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 (expression array or RNA-Seq) datasets, and DNA promoter occupancy in cistromic (ChIP-Seq) datasets. Conventionally, only a small fraction of data points from such datasets are characterized in any level of detail in the associated hypothesis-driven articles. Although the remaining 'omics data points possess potential collective re-use value for validating experimental data or gathering evidence to model cellular signaling pathways, the findability, accessibility, interoperability and re-use (FAIR) status of these datasets^{3,4} has been historically limited.

¹Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, 77030, USA. ²Duncan NCI Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, 77030, USA. ³Icahn School of Medicine, Mount Sinai University, New York, NY, 10029, USA. ⁴Department of Pharmacology and Cancer Biology, Duke University School of Medicine, Durham, NC, 27710, USA. ⁵Diabetes & Metabolism Research Institute, City of Hope, Duarte, CA, 91010, USA. ⁶Department of Chemical Physiology, Scripps Research Institute, La Jolla, CA, 92037, USA. ⁷University of Michigan School of Public Health, Ann Arbor, MI, 48109, USA. ⁸These authors contributed equally: David Abraham and Kirt Martin. *email: nmckenna@bcm.edu

We previously described our efforts to biocurate transcriptomic datasets involving genetic or small molecule manipulation of nuclear receptors⁵. Here we describe here a novel and distinct web knowledgebase, the Signaling Pathways Project (SPP), that enhances the FAIR status of public cell signaling ‘omics datasets along three dimensions. Firstly, SPP encompasses datasets involving genetic and small molecule perturbations of a broad range of cellular signaling pathway modules - receptors, enzymes, transcription factors and their co-nodes. Secondly, SPP integrates transcriptomic datasets with biocurated ChIP-Seq datasets, documenting genomic occupancy by transcription factors, enzymes and other factors. Thirdly, we have developed a meta-analysis technique that surveys across transcriptomic datasets to generate consensus ranked signatures, referred to as consensomes, which allow for prediction of signaling pathway node-target regulatory relationships. We have validated the consensomes using alignment with literature knowledge, integration of transcriptomic and ChIP-Seq evidence, and using bench experimental use cases that corroborate signaling pathway node-target regulatory relationships predicted by the consensomes. Finally, we have designed a user interface that makes the entire data matrix available for routine data browsing, mining and hypothesis generation by the mammalian cell signaling research community at <https://www.signalingpathways.org>.

Results

SPP overview. Mammalian signal transduction pathways comprise four major categories of pathway module: activated transmembrane or intracellular receptors, which initiate the signals; intracellular enzymes, which propagate and modulate the signals; transcription factors, which give effect to the signals through regulation of gene expression; and co-nodes, a broad variety of molecular classes, such as many transcriptional coregulators⁶, that do not fall into the other three categories. Figure 1 shows the scope of the SPP knowledgebase in terms of the major signaling pathway module categories, classes and node families, as well as the biosample classification of tissues and cell lines in which these nodes are studied. Table 1 shows representative examples of the hierarchical relationships within each of the SPP signaling pathway module categories. Having defined relationships within each major signaling pathway module, we proceeded to develop a dataset biocuration pipeline (Fig. 2) 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, as well as their biosample of study. To make the results of our biocuration efforts routinely and freely available to the research community, we next developed a web user interface (UI) 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 available in Supplementary Information Section 1.

Browsing of SPP datasets. The full dataset listing (<https://www.signalingpathways.org/datasets/index.jsf>; see also Supplementary Information Section 1A for a user walk-through) can be filtered using any combination of: ‘omics dataset type; SPP category (receptor, enzyme, transcription factor, co-node); class or family; biosample physiological system and organ; or species. Individual dataset pages enable integration of SPP with the research literature via digital object identifier (DOI)-driven links from external sites, as well as for citation of datasets to enhance their FAIR status^{3,4}.

Mining of SPP datasets in Ominer. The SPP query interface, Ominer, allows a user to specify single gene target, Gene Ontology (GO) term or 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. 3a). Single gene queries are designed for researchers who wish to evaluate transcriptomic or ChIP-Seq evidence for regulation of a single gene of interest across all nodes, or within specific categories, classes or families (see Table 2 for examples and Supplementary Information Section 1B for a user walk-through). GO term queries (see Table 3 for examples and Supplementary Information Section 1C for a user walk-through) accommodate users interested not in a specific gene, but rather in regulation of multiple genes mapped to broader functional or mechanistic terms by GO annotators. Gene list queries (see Supplementary Information Section 1D for a user walk-through) return transcriptional regulatory data points for custom user gene lists containing up to 500 approved gene symbols or Entrez GeneIDs.

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 genomic target(s) of interest (Fig. 3b, transcriptomic & 3c, cistromic/ChIP-Seq). The vertical organization of the default Category view in both transcriptomic and cistromic/ChIP-Seq Regulation Reports reflects conventional schematic depictions of cellular signaling pathways, with Receptors on top, followed by Enzymes, Transcription Factors and Co-nodes. Reflecting the hierarchy in Table 1, each Regulation Report category is subdivided into classes (depicted as **Category|Class** in the UI) which are in turn subdivided into families containing member nodes, which are themselves mapped to bioactive small molecules (BSMs) that regulate their function. The transcriptomic Regulation Report (Fig. 3b) 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) manipulations of nodes within a given family. Below the node sections, the transcriptomic Regulation Report contains a Models section, in which data points from related animal and cell model experiments are consolidated to convey evidence for previously underappreciated roles of a target transcript in specific physiological contexts, such as adipogenesis. 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. The cistromics/ChIP-Seq Report (Fig. 3c) displays ChIP-Atlas⁷-generated 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. 3c). 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” or “Species” views from the dropdown, as shown in Fig. 3b.

| Signaling Pathway Module | | | | Bioactive small molecule (BSM) | | | |
|--------------------------|-----------------------|----------------------|----------|--------------------------------|---------|-------------------|-------------------|
| Category | Class | Family | Node | Ligand | Drug | Synthetic organic | Natural substance |
| Receptors | G protein-coupled | Adrenoreceptors | ADRB1 | EPI | ISOPREN | — | — |
| | Catalytic | EGF receptors | EGFR | EGF | GEFIT | WZ002 | — |
| | Nuclear | PPARs | PPARG | LINO | ROSI | GW544 | — |
| Enzymes | Deacetylases | Histone deacetylases | HDAC3 | — | VORIN | — | BUTYR |
| | Kinases | Abl family kinases | ABL1 | — | IMAT | — | ZINC498 |
| | Cyclases | Adenylyl cyclases | ADCY3 | — | — | — | FORSK |
| Transcription Factors | Basic leucine zipper | CREB-like factors | CREB1 | — | — | — | — |
| | Forkhead/winged helix | FOXO | FOXO1 | — | — | — | THSP |
| Co-nodes | Heat shock proteins | HSP 90 proteins | HSP90AA1 | — | — | — | GEDUN |

Table 1. Examples of signaling pathway module hierarchies in the SPP knowledgebase. Nodes and peptide BSMs are represented by approved HGNC symbol for the encoding gene. BSMs are abbreviated as follows: BUTYR, butyric acid; EPI, epinephrine; FORSK, forskolin; GEDUN, gedunin; GEFIT, gefitinib; GW544, GW409544; IMAT, imatinib; ISOPREN, isoprenaline; LINO, linoleic acid; ROSI, rosiglitazone; THSP, thrombospondin; VORIN, vorinostat; WZ002, WZ4002; ZINC498, ZINC08764498.

manuscripts or grant applications, all Reports are accessible by a constructed URL defining all of the individual query parameters.

Consensomes: discovering downstream genomic targets of signaling pathway nodes. 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 genomic targets. Accordingly, we next set out to design a meta-analysis pipeline that would leverage our biocuration platform to reliably rank signaling pathway node-target gene regulatory relationships in a given biosample context. Since this analysis was designed to establish a consensus for a node or node family across distinct datasets from different laboratories, we referred to the resulting node-target rankings as consensomes. A detailed description of the biocuration and statistical methodologies behind transcriptomic and cistromic/ChIP-Seq consensome analysis is provided in the Methods section.

Consensome queries (see Supplementary Information Subsection 1E for a walk-through) are designed for users unfamiliar with a particular signaling node family who are seeking evidence for targets that have close regulatory relationships with members of that family. Table 4 shows examples of the consensomes available in the initial version of the SPP knowledgebase. Section 2 of the Supplementary information shows the full list of consensomes available in the initial release of SPP. Consensomes are 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 (see Supplementary Information Subsection 1E). Subsequent menus allow for selection of specific signaling pathway node families, physiological systems or organs of interest, or species. To accommodate researchers interested in a specific physiological system or organ rather than a specific pathway node, consensomes are also calculated across all experiments mapping to a given physiological system (metabolic, skeletal, etc.) and organ (liver, adipose tissue, etc.), providing for identification of targets under the control of a broad spectrum of pathway nodes in those organs. To maximize their distribution, exposure and citation in third party resources, consensomes can also be accessed by direct DOI-resolved links.

Consensomes are displayed in an accessible tabular format (Fig. 4) in which the default ranking is in ascending order of consensome p-value (CPV; see Methods), although targets can be ranked by any column desired. To reflect the frequency of differential expression of a given target relative to others in a specific 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 genomic 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 variable in consensome analysis, the direction of differential expression is omitted when calculating the ranked signatures. That said, an appreciation of the pharmacology of a specific node-target gene relationship is essential to allow researchers to place the ranking in a specific biological context and to design subsequent experiments in an informed manner. To accommodate this, consensome targets link to transcriptomic or cistromic Regulation Reports filtered to display those data points (i.e. all $p < 0.05$ fold changes) that contributed to the calculation of the CPV for each target.

Validation of consensomes. We next wished to verify that consensomes were reliable reference datasets for modeling regulatory relationships between cellular signaling pathway nodes and their downstream genomic targets. To do this we designed a validation strategy comprising four components: comparison of consensomes with existing canonical (i.e. literature-defined) node-target relationships; reciprocal validation of node-target

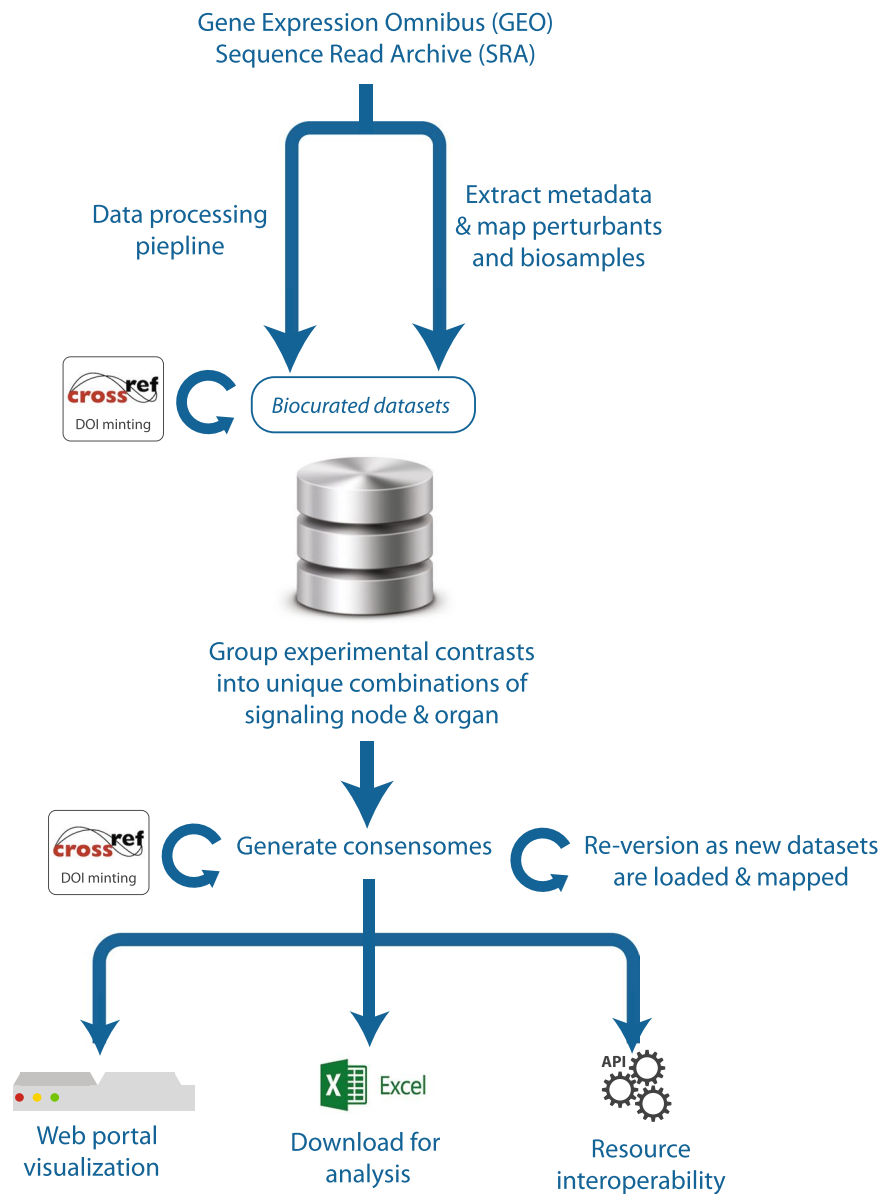


Fig. 2 Schematic depiction of SPP biocuration and FAIR annotation pipeline. See the Methods section for additional information.

relationships between transcriptomic and ChIP-Seq consensomes; validation of pan-node organ consensomes; and in three experimental use cases that functionally validate predicted node-target relationships.

Canonical signaling node downstream targets are highly ranked in transcriptomic and cistromic nuclear receptor consensomes.

Two considerations recommended members of the nuclear receptor (NR) superfamily of physiological ligand-regulated transcription factors for selection for initial 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 of their downstream targets. 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. We selected the ten top ranked targets in the following consensomes: estrogen receptors in human mammary gland (ERs-Hs-MG); the androgen receptor in human prostate gland (AR-Hs-Prostate); the glucocorticoid receptor in mouse liver (GR-Mm-Liver); and the peroxisome proliferator-activated receptor (PPAR) family in the mouse metabolic system (PPARs-Mm-Metabolic). Encouragingly, we found that 36/40 (90%) of the most highly ranked targets across all four consensomes had been previously identified as targets of members of those node families in the research literature, and that of these same 40 genes, 82% (33/40) were in the 90th percentile or higher in the corresponding ChIP-Seq consensomes (Supplementary Information Section 3).

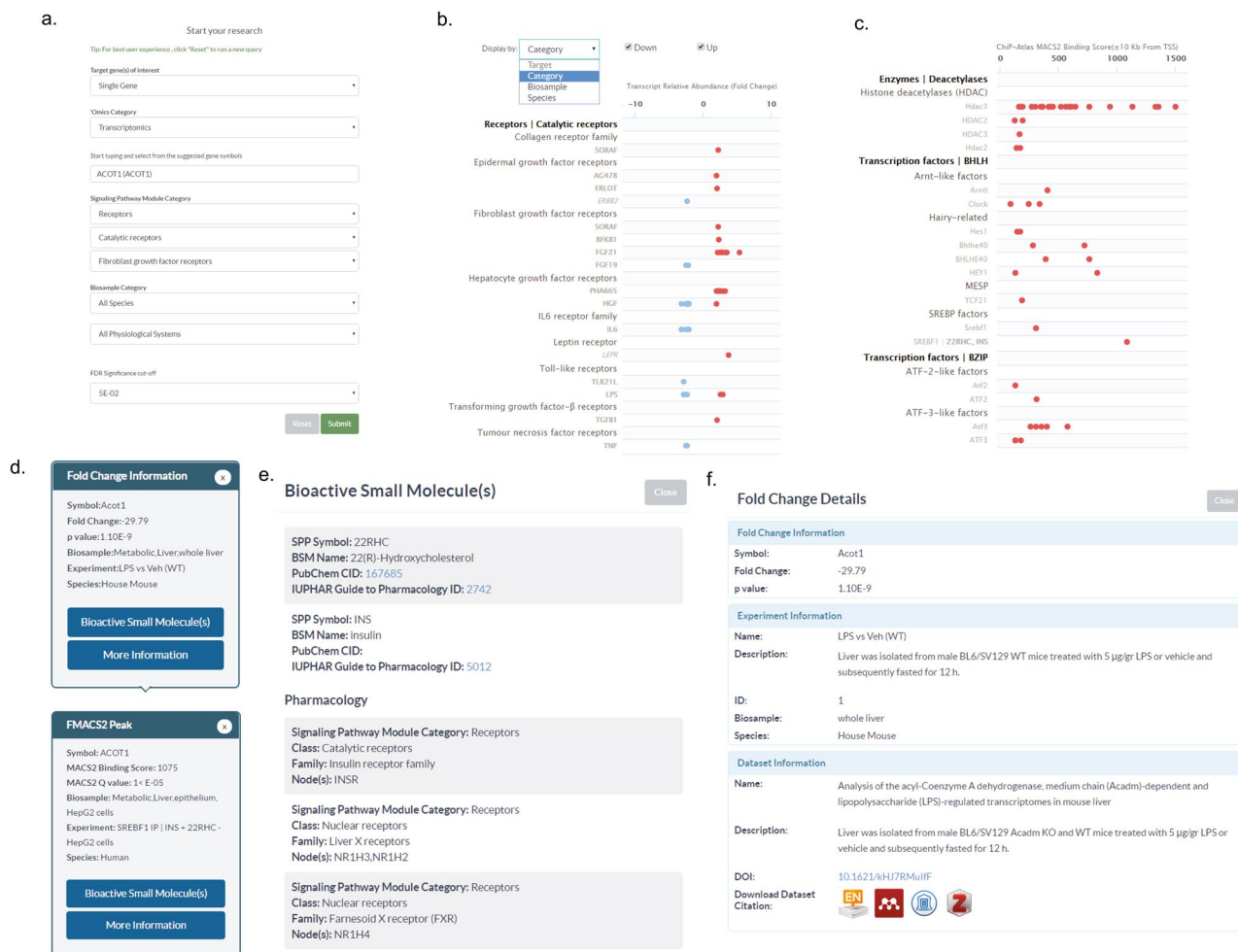


Fig. 3 Key elements of the SPP query and reporting interface. **(a)** Ominer query form. **(b)** The transcriptomic Regulation Report. The default display for single gene queries is by Category, which can be adjusted to cluster data points by biosample or species. The default display for multi-gene queries 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.

Frequently regulated hepatic transcripts are enriched for critical regulators of metabolic pathways. Substantial literature evidence from prokaryotic⁸ and eukaryotic⁹ systems indicates that genes encoding metabolic enzymes are transcriptionally plastic and subject to dynamic regulation of their expression by numerous afferent metabolic and endocrine cues. If consensome analysis were biologically valid, we anticipated that targets with elevated rankings in the murine hepatic transcriptomic consensome – that is, genes that are preferentially responsive to multiple hepatic signaling pathways – would be enriched for genes encoding enzymes with prominent roles in hepatic metabolism. To test this hypothesis, we first identified genes in the 99th percentile of the All nodes-Mm-liver transcriptomic consensome: that is, the top 1% of genes that are significantly differentially expressed in expression profiling experiments in a murine hepatic biosample, irrespective of the experimental design (Fig. 5, orange data points, $n = 258$; Supplementary Information Section 4; <http://tiny.cc/guqk8y>). Based upon a set of 1647 murine metabolic enzymes curated by the Mammalian Metabolic Enzyme Database resource (Corcoran, 2017 #250) we found that 38% (99/258) of the top 1% genes encoded metabolic enzymes, a 6.5-fold enrichment over the frequency of metabolic enzyme-encoding genes in the entire All nodes-Mm-liver transcriptomic consensome (5.7%, 1490/25922).

We next speculated that transcripts under fine control by hepatic signaling pathways would be enriched for enzymes whose deficiency would have a critical impact upon hepatic metabolism. Using the OMIM resource¹⁰, we identified a set of unique human genes whose deficiency has a literature-supported connection to a human metabolic disease or trait ($n = 5277$). Using this reference gene set, we established that human orthologs of 40% (41/99) of metabolic enzyme-encoding targets in the 99th percentile of the All Nodes-Mm-liver transcriptomic

| 'Omics Category | Pathway Module (Category, Class, Family) | Biosample (System, Organ) | Target | SPP Query URL |
|----------------------|---|---------------------------|----------------|---|
| Transcriptomic | All | All | <i>PDK4</i> | http://tiny.cc/pdk4tx |
| | All | Metabolic, Liver | <i>ALDH3A2</i> | http://tiny.cc/ujvk8y |
| | Receptors, Catalytic, Insulin receptor | All | <i>PMAIP1</i> | http://tiny.cc/kuvk8y |
| | Enzymes, Kinases, Cyclin-dependent | All | <i>GINS3</i> | http://tiny.cc/0yvk8y |
| Cistromic (ChIP-Seq) | All | All | <i>LHPP</i> | http://tiny.cc/lhppcx |
| | All | Male Repro, Prostate | <i>TMPRSS2</i> | http://tiny.cc/v9vk8y |
| | Enzymes, Acetyltransferase, CBP/p300 | All | <i>MAMDC2</i> | http://tiny.cc/vfwk8y |
| | Transcription factors, BZIP, C/EBP family | Immune, Leukocytes | <i>TRIB1</i> | http://tiny.cc/eiwk8y |

Table 2. Examples of Single Gene queries in the SPP 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 more stringent differential expression or significance cut-offs. Please refer to Supplementary Information Section 1B for a guide to constructing Single Gene queries.

| 'Omics Category | Pathway Module (Category, Class, Family) | GO Term | SPP Query URL |
|----------------------|--|--|---|
| Transcriptomic | Receptors, Catalytic, All | Fatty acid beta oxidation | http://tiny.cc/rbyk8y |
| | Receptors, G-protein coupled, All | Glycolytic process | http://tiny.cc/kdyk8y |
| | Enzymes, Kinases, Cyclin-dependent | Adipose tissue development | http://tiny.cc/whyk8y |
| | Enzymes, Acetyltransferases, All | Acute inflammatory response | http://tiny.cc/8izk8y |
| Cistromic (ChIP-Seq) | Receptors, Nuclear, PPARs | Cellular response to fatty acid | http://tiny.cc/zrzk8y |
| | Enzymes, E3 Ubiquitin ligases, All | Immune response | http://tiny.cc/s8zk8y |
| | Transcription factors, p53 domain, All | Cellular response to DNA damage stimulus | http://tiny.cc/hnzk8y |
| | Transcription factors, BZIP, All | Urea cycle | http://tiny.cc/49zk8y |

Table 3. Examples of GO Term queries in the SPP knowledgebase. All example queries are for human gene annotations. GO Term queries must specify at least a Pathway Module Class. Please refer to Supplementary Information Section 1C for a guide to constructing GO term queries.

consensome had documented deficiencies in a human metabolic disorder (Table 5 and Supplementary Information Section 5), compared with a frequency of 15% (3865/25922) for such genes in the entire consensome.

Rate-limiting enzymes (RLEs) play critical roles in determining mammalian metabolic flux¹¹. We surmised that hepatic signaling pathways might preferentially target RLE-encoding genes to exert efficient control over hepatic metabolism, and that this would be reflected in the enrichment of RLE genes among enzyme-encoding targets in the top 1% of the All Nodes-Mm-liver transcriptomic consensome. Consistent with this notion, 40/99 (40%) of the metabolic enzymes encoded by targets in this group of genes catalyze rate-limiting steps in the pathways in which they participate (Supplementary Information Sections 4 and 5), a nearly 3-fold enrichment over the previously published estimate of 14% (96/687) for the proportion of hepatic metabolic enzymes made up by RLEs¹¹. Collectively, these analysis demonstrates the ability of pan-node organ consensomes to illuminate factors that are downstream targets of multiple distinct signaling nodes in a specific organ and, by inference, have pivotal, tightly-regulated roles in the function of that organ.

We next wished to establish whether the biological significance implied by elevated rankings in consensomes for cellular signaling pathway node families was reflected in both gain- and loss-of-function validation experiments at the bench.

Bench validation use case 1: elevated consensome rankings predict functional roles for targets in signaling node pathways.

Figure 6a shows a scatterplot depiction of the ERs-Hs-All organs consensome. The two distinct tails in the distribution demarcate between genes whose discovery rates are comparable, but based upon different total numbers of experiments in which the genes were assayable, and therefore giving rise to different CPVs. We used Q-PCR analysis to verify ER-dependent regulation of a panel of both characterized and uncharacterized ER targets (highlighted in orange in Fig. 6a) that were highly ranked in the ERs-Hs-All consensome (Fig. 6b). We next identified targets that were assigned very high consensome rankings, but whose functional importance in the context of signaling by the corresponding signaling nodes has been largely uncharacterized in the research literature. 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 cancers¹². Despite a ranking in the transcriptomic (ERs-Hs-All-TC CPV = <1E-130, 99.99th 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. 6b, *TPD52L1*), the functional role of *TPD52L1* in ER signaling has gone unexplored in the research literature. Suggestive of a role for *TPD52L1* in ER regulation of cell division, we identified 17BE2-dependent association of *TPD52L1* with structures resembling stress fibers

| Omics Category | Pathway Module (Category, Class, Family) | Biosample (System, Organ, Species) | SPP Query URL |
|----------------|---|------------------------------------|---|
| Transcriptomic | Receptors, Catalytic, EGF receptors | Female Rep, Mammary gland, Hs | http://tiny.cc/5g0k8y |
| | Receptors, Nuclear, PPARs | Metabolic, Liver, Mm | http://tiny.cc/al0k8y |
| | Enzymes, Kinases, Cyclin-dependent | Female Rep, All, Hs | http://tiny.cc/ep0k8y |
| | Co-nodes, RNA binding domain, PPARGC1 coactivator 1 (PPARGC1) | Metabolic, All, Mm | http://tiny.cc/8s0k8y |
| | Transcription factors, SAND domain, AIRE | Immune, Thymus, Mm | http://tiny.cc/6w0k8y |
| | All | Metabolic, Liver, Mm | http://tiny.cc/guqk8y |
| | All | Metabolic, Adipose, Mm | http://tiny.cc/f8qk8y |
| ChIP-Seq | Enzymes, Acetyltransferases, CBP/p300 | All, All, Hs | http://tiny.cc/1d1k8y |
| | Enzymes, Acetyltransferases, NCOA family | All, All, Hs | http://tiny.cc/we1k8y |
| | Transcription factors, BZIP, C/EBP family | All, All, Mm | http://tiny.cc/xh1k8y |
| | Transcription factors, E2F/FOX, E2F family | All, All, Hs | http://tiny.cc/tj1k8y |
| | Transcription factors, E2F/FOX, FOXO family | All, All, Mm | http://tiny.cc/jn1k8y |

Table 4. Examples of consensomes in the SPP knowledgebase. Consensomes are calculated at either: the signaling pathway node family level, ranking targets based on their transcriptional sensitivity to manipulation of nodes in a given gene family; or across all experiments in a given organ biosample, to indicate frequently regulated targets in a given organ. Female Rep, Female reproductive; Hs, human; Mm, mouse. Please refer to Supplementary Information Section 1D for a guide to constructing consensome queries.

(Fig. 6c), which play an important role in mitosis orientation during cell division¹³, and found that depletion of *TPD52L1* resulted in a significant decrease in 17BE2-induced proliferation of MCF-7 cells (Fig. 6d).

MBOAT2 encodes an enzyme catalyzing cycles of glycerophospholipid deacylation and reacylation to modulate plasma membrane phospholipid asymmetry and diversity¹⁴. *MBOAT2* has rankings in both the AR-Hs-All TC (CPV = 3.7E-38, 99.96th percentile) and ChIP-Seq (99th percentile) consensomes comparable to those of canonical AR target genes such as *KLK3* and *TMPRSS2*. In contrast to the large volume of literature devoted to these targets however, with the exception of a mention in a couple of androgen expression profiling studies^{15,16}, the functional role of *MBOAT2* in the context of AR signaling has been entirely unstudied. In validation of its elevated AR consensome ranking, we confirmed that *MBOAT2* was an AR-regulated gene in cultured prostate cancer cell lines (Fig. 6e), and that depletion of *MBOAT2* significantly increased LNCaP cell numbers at growth day 5 in in R1881-treated cells, but not untreated cells (Fig. 6f). 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 epithelium¹⁷. Such an assertion is supported by the recent characterization of the role of *MBOAT2* in chondrogenic differentiation of ATDC5 cells¹⁸, and by the fact that the AR agonist testosterone stimulates the chondrogenic potential of chondrogenic progenitor cells¹⁹.

Bench validation use case 2: GR, ERR family members and insulin receptor regulate targets encoding glycogen synthase phosphatase and kinase regulatory subunits.

The experimental validation studies in the first use case 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. Glycogen synthase, which catalyzes the rate-limiting step in the interconversion of glucose and glycogen in metabolic organs, is subject to tandem activation by protein phosphatase 1 (PP1)²⁰, and deactivation by 5'AMP-activated protein kinase (AMPK)²¹. Although regulation of glycogen metabolism in a variety of organs is known to be under the control of signaling mediated by the glucocorticoid (GR)²², estrogen receptor-related (ERR)²³ and insulin (IR)²⁴ receptor families, the respective underlying mechanisms are incompletely understood. We wished to use SPP consensomes to investigate the hypothesis that regulation of glycogen metabolism by members of these distinct receptor families might involve convergent regulation of glycogen synthase activity. Surveying $p < 0.05$ genes in the mouse ERR and GR and human IR transcriptomic consensomes mapping to the GO terms “AMP-activated protein kinase activity” or “protein phosphatase regulator activity”, we isolated two targets, *Ppp1r3c* (GR-Mm-All CPV = 1.89E-15; ERR-Mm-All CPV 5.98E-06) and *Prkab2* (ERR-Mm-All CPV = 1.84E-04, IR-Hs-All CPV = 4.9E-04). These encode, respectively, the PTG regulatory subunit of the PP1 holoenzyme^{25,26}, and the AMPK β 2 regulatory subunit of the AMPK holoenzyme^{27,28}. Corroborating these predicted regulatory relationships, we identified conserved GR and ERR response elements in the *Ppp1r3c* promoter (Supplementary Information Section 6). We also confirmed that *Ppp1r3c* was a target of GR in mouse Hepa-1-c liver cells (Fig. 7a) and that both *Ppp1r3c* and *Prkab2* were transcriptionally regulated in response to either loss- or gain-of function of *Esrra* in skeletal muscle (Fig. 7b–d, left panel). Finally, consistent with the recently-documented regulation of AMPK activity by IGF1 signaling²⁹, treatment of myoblasts for 24 h with the IR agonist IGF1 stimulated the *Prkab2* promoter, an effect that was further enhanced by expression of both *Esrra* and *Esrrg* isoforms (Fig. 7d, right panel).

Bench validation use case 3: the murine ERR, PPARGC and adipose tissue consensomes implicate *Mcrip2* in adipocyte oxidative metabolism.

The control of cellular mitochondrial content and

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 entries

Search:

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

Fig. 4 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 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.

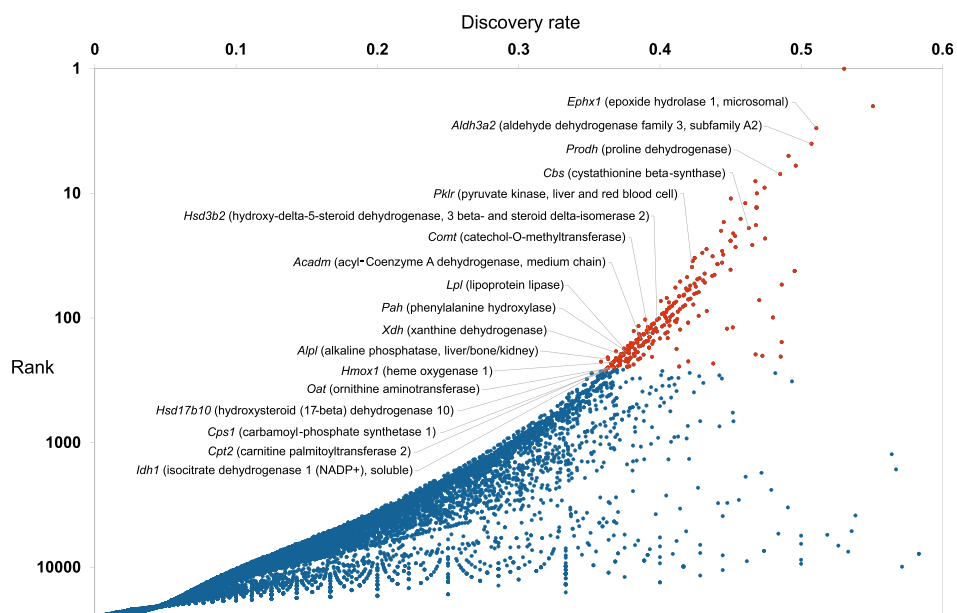


Fig. 5 Scatterplot of the mouse all nodes liver transcriptomic consensome. This plot distills data from nearly 300 distinct experiments to convey a visual appreciation of the relative rates of differential expression of murine genes across a variety of hepatic signaling contexts. Genes in the 99th percentile are highlighted in orange. For cross-reference with Table 5, genes encoding selected metabolic enzymes in the 99th percentile are called out by gene symbol and name. For details, refer to the Transcriptomic Consensomes subsection in the Generation of Consensomes section of the Methods.

| Target | SPP query URL | CPV | Hepatic metabolic pathway | Known human deficiency disease |
|---------------------------------|---|-----------|--|----------------------------------|
| Lipid metabolism | | | | |
| <i>Ephx1</i> | http://tiny.cc/m61k8y | 7.27E-107 | Conversion of epoxides to trans-dihydrodiols | Familial hypercholelanemia |
| <i>Aldh3a2</i> | http://tiny.cc/6b2k8y | 2.11E-103 | Oxidation of fatty aldehydes to fatty acids (RL) | Sjogren-Larsson syndrome |
| <i>Acadm</i> | http://tiny.cc/kf2k8y | 9.52E-65 | Medium-chain fatty acid β -oxidation (RL) | ACADM deficiency |
| <i>Lpl</i> | http://tiny.cc/gj2k8y | 1.44E-62 | Hydrolysis of TGs from TG-rich lipoproteins (RL) | Familial combined hyperlipidemia |
| <i>Cpt2</i> | http://tiny.cc/2l2k8y | 3.22E-58 | Mitochondrial fatty acid β -oxidation (RL) | Neonatal CPT II deficiency |
| Carbohydrate metabolism | | | | |
| <i>Pfkr</i> | http://tiny.cc/5n2k8y | 1.04E-79 | Glycolysis (RL) | Pyruvate kinase deficiency |
| <i>Idh1</i> | http://tiny.cc/7p2k8y | 3.21E-58 | TCA cycle (RL) | Glass syndrome, Ollier disease |
| Amino acid metabolism | | | | |
| <i>Prodh</i> | http://tiny.cc/ws2k8y | 2.54E-94 | Oxidation of proline to glutamate (RL) | Hyperprolinemia |
| <i>Pah</i> | http://tiny.cc/602k8y | 3.09E-62 | Phenylalanine catabolism (RL) | Phenylketonuria |
| <i>Cth</i> | http://tiny.cc/c42k8y | 9.20E-59 | Cysteine synthesis (RL) | Cystathioninuria |
| <i>Oat</i> | http://tiny.cc/752k8y | 3.14E-58 | Glutamate biosynthesis | Gyrate atrophy |
| Other metabolic pathways | | | | |
| <i>Cbs</i> | http://tiny.cc/h92k8y | 6.47E-87 | Trans-sulfuration pathway (RL). | Homocysteinemia |
| <i>Hsd3b2</i> | http://tiny.cc/4f3k8y | 1.01E-68 | Aldosterone biosynthesis (RL) | Congenital adrenal hyperplasia |
| <i>Comt</i> | http://tiny.cc/1g3k8y | 5.26E-65 | Degradation of catecholamines | Panic disorder, schizophrenia |
| <i>Alpl</i> | http://tiny.cc/cp4k8y | 1.86E-60 | General hydrolysis of phosphate esters | Adult hypophosphatasia |
| <i>Xdh</i> | http://tiny.cc/mq4k8y | 2.33E-61 | Purine metabolism (RL) | Type I xanthinuria |
| <i>Hmox1</i> | http://tiny.cc/pv4k8y | 2.93E-59 | Heme degradation (RL) | HMOX1 deficiency |
| <i>Cps1</i> | http://tiny.cc/ww4k8y | 3.14E-58 | Mitochondrial urea cycle (RL) | CPS1 deficiency |
| <i>Hsd17b10</i> | http://tiny.cc/h14k8y | 3.21E-58 | 17 β -oxidation of androgens and estrogens | HSD10 mitochondrial disease |

Table 5. Selected genes encoding metabolic enzymes in the 99th percentile of the All nodes-Mm-liver transcriptomic consensome & deficiency in whose human orthologs is associated with a metabolic disorder. SPP Query URL links point to transcriptomic Regulation Reports filtered for mouse liver ($FC > 1$ & $p < 0.05$). Gene names corresponding to gene symbols are shown in Fig. 5. See Supplementary Information Section 5 for the full list. RL, rate-limiting reaction.

oxidative capacity is important for cellular and organismal energy homeostasis³⁰. For example, brown and beige adipocytes generate new mitochondria and increase their oxidative and thermogenic capacity in response to norepinephrine (NE), which is secreted locally when the organism senses a cold environment³¹. NE (adrenergic) stimulation elicits an acute transcriptional response, exemplified by the induction of genes such as the uncoupling protein *Ucp1*, the Pparg co-node *Ppargc1a* and the signaling regulator *Gadd45g*^{31,32}. *In vivo*, chronic or repeated exposure to cold (or to adrenergic agonists) also leads to higher mitochondrial DNA content, increased cristae density and enhanced expression of oxidative enzymes (OxPhos complexes) and *Ucp1*³¹.

Mitochondrial biogenesis is regulated by a variety of nuclear receptors, including members of the ERR family, as well as Pparg and its co-nodes *Ppargc1* and *Ppargc1b*, members of the PPARGC family of RNA-binding transcriptional coregulator co-nodes³³. A highly ranked gene in the ERRs-Mm-All transcriptomic consensome was *Mcrip2* (CPV = 1.54E-12), which has no literature-characterized function or role other than a report identifying it as an interacting partner of Ddx6 that was localized to RNA stress granules³⁴. Interestingly, we noted that *Mcrip2* was also very highly ranked in the All nodes-Mm-adipose (<http://tiny.cc/f8qk8y>; CPV = 6.58E-37), PPARs-Mm-Adipose (<http://tiny.cc/rbrk8y>; CPV = 1.17E-30), PPARGC1s-Mm-Metabolic (<http://tiny.cc/7drk8y>; CPV = 1.84E-04) and All nodes-Mm-liver (<http://tiny.cc/guqk8y>; CPV = 2.2E-88) transcriptomic consensomes, indicating potentially influential roles in adipose and hepatic biology. Inspection of the *Mcrip2* transcriptomic (<http://tiny.cc/mcrip2tx>) and cistromic (<http://tiny.cc/mcrip2cx>) Regulation Reports indicated that it was regulated under conditions of mitochondrial biogenesis in adipose tissue, as well as loss and gain of function of a variety of known signaling node regulators of mitochondrial biogenesis. Corroborating this evidence, we identified a conserved consensus ERR binding site in the first intron of *Mcrip2* (Supplementary Information Section 7) and confirmed that *Esrra* and *Ppargc1a* were recruited to the *Mcrip2* ERRE (Fig. 7e). We also confirmed the interdependence of *Esrra*, *Ppargc1a* and *Ppargc1b* in regulation of *Mcrip2* in mouse muscle cells (Fig. 7f; Supplementary Information Section 7) and brown adipocytes (Fig. 7f), and demonstrated induction of *Mcrip2* in response to conditions mimicking chronic adrenergic stimulation of brown adipocytes (Fig. 7g).

Discussion

Effective re-use of 'omics datasets in the field of cellular signaling relies upon the ability of bench researchers to ask sophisticated questions across this universe of data points in a routine manner. Many excellent tools and resources have been developed in the field of cell signaling 'omics³⁵⁻⁴⁶. Here, we set out to complement these resources to allow researchers to routinely answer targeted questions such as: what cell cycle-related factors are regulated by FGF receptors in human liver? What genomic targets are most responsive to insulin receptor

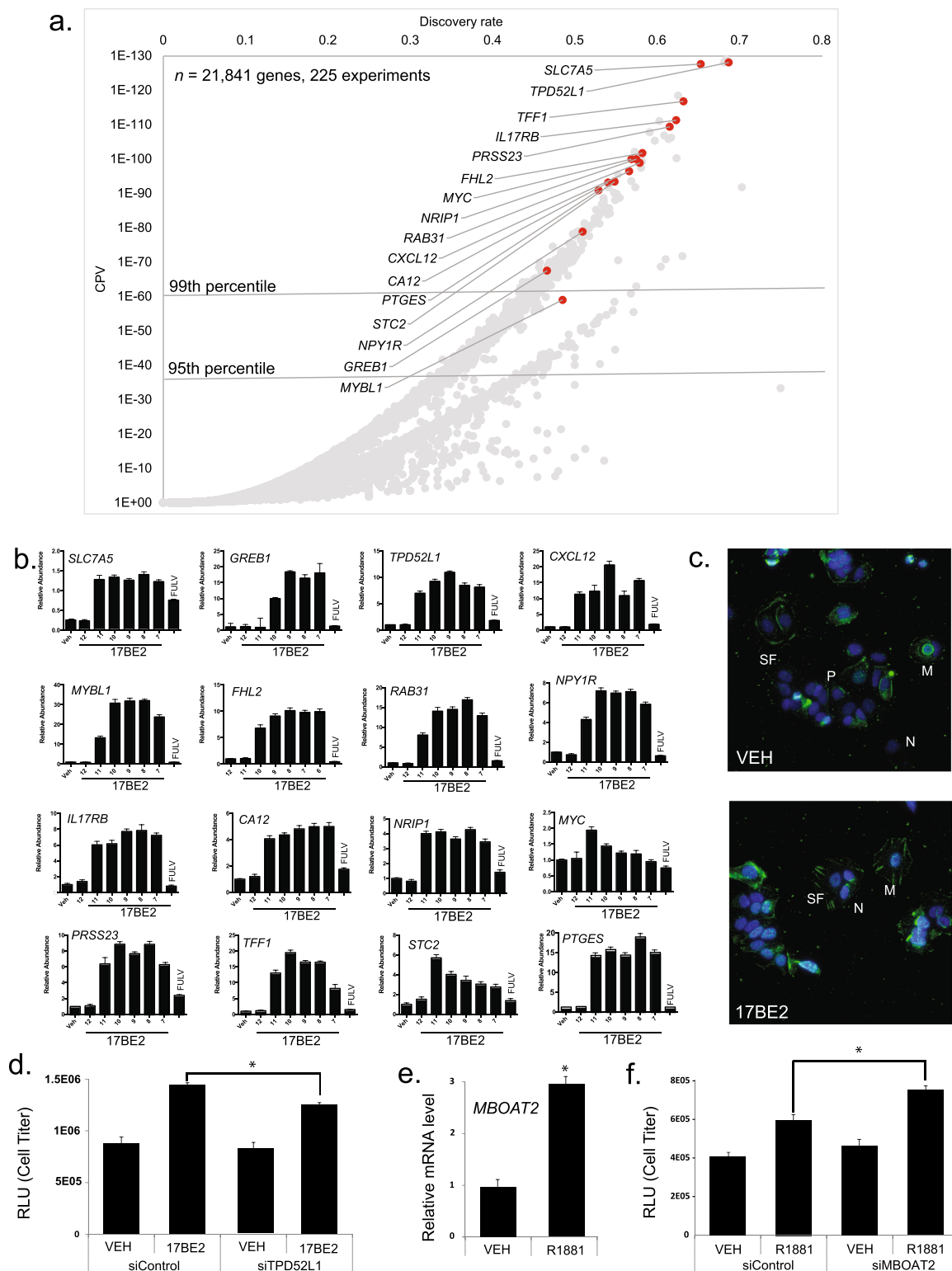


Fig. 6 Validation of ER regulation of *TPD52L1* and AR regulation of *MBOAT2*. **(a)** Scatterplot of the ERs-Hs-MG transcriptomic consensome. Genes selected for Q-PCR validation are colored orange and called out by approved gene symbol. 99th and 95th percentile cut-offs are shown for reference. **(b)** Q-PCR analysis of dose dependent induction by 17BE2 in MCF-7 cells of targets with elevated rankings in the ERs-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 of the selective ER downregulator FULV. Consistent with the strong ER family node dependence of regulation predicted by the ERs-Hs-MG transcriptomic and ChIP-Seq consensomes, FULV completely abolishes 17BE2 induction of all target genes tested. 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. **(c)** 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 10 μm . (in the inset, 5 μm). M, membrane; N, nucleus; P, perinuclear junctions; SF, stress fibers. **(d)** Depletion of TPD52L1 restricts MCF-7 cell viability. **(e)** Induction of *MBOAT2* in LNCaP prostate epithelial cells upon treatment with 0.1 nM AR agonist R1881. **(f)** 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. * $p < 1\text{E-}04$.

signaling in the liver? What targets in my gene set are regulated by E3 ubiquitin ligases? To fill this gap, we designed a knowledgebase, SPP, which allows bench researchers to routinely evaluate evidence in public transcriptomic or ChIP-Seq datasets to infer the roles of cellular signaling pathway nodes in their system of interest.

The SPP resource is characterized by a number of unique features. Previous transcriptomic meta-analysis approaches in the field of cellular signaling have been perturbation-centric, and applied to experiments involving a single unique perturbant^{47,48}. 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 accordingly incorporates a step that maps experiments to their relevant pathway nodes. This mapping step affords the consensomic analysis greater statistical power, enabling it to call potential node-target relationships with a higher degree of confidence than would otherwise be possible. Incorporation of this mapping approach into the Regulation Reports also serves to place emphasis on the functional relatedness of distinct experimental perturbations with respect to a given node-target regulatory relationship. An additional unique aspect is that many 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 to our knowledge the first meta-analysis to be sustainably integrated into an actively-biocurated FAIR web resource in a manner supporting routine dataset re-use and citation by bench researchers lacking formal informatics training.

Our resource has a number of limitations. SPP is currently based upon transcriptomic and ChIP-Seq data, since these are the most numerous and informatically mature of the various types of 'omics data. Future versions of the knowledgebase will only benefit from the incorporation of the growing number of metabolomic and proteomic profiling datasets, which will illuminate effects of signaling pathways on cellular functions not addressed by transcriptional methodologies. Secondly, bias in publically archived datasets towards specific nodes and biosamples is to some extent reflected in SPP. 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, or in a specific organ or tissue) and if such experiments do not exist or are otherwise publically unavailable, these targets would not 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 (e.g. specific organ) for such regulation to take place. Caveats such as these notwithstanding, we believe SPP adds value to the currently available tool set enabling bench investigators to re-use archived discovery scale transcriptional datasets for hypothesis generation and data validation.

Methods

Data model design. The goal of 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 expression⁴⁹. 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 adapting existing, mature classifications for receptors (International Union of Pharmacology, IUPHAR⁵⁰), enzymes (International Union of Biochemistry and Molecular Biology Enzyme Committee⁵¹) and transcription factors (TFClass⁵²). Molecular classes that are relevant to cellular signaling pathways but do not fall into any of the three categories referred to above, such as regulatory RNAs, chromatin factors and cytoskeletal components, were assigned to a "co-nodes" category, classified according to approved genome nomenclature committee-defined gene families. Table 1 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 node families, to which individual node genes were assigned. Figure 1a–c summarizes the scope of the major classes and/or families in each category, collectively comprising 174 families of receptors, 616 families of enzymes and 371 families of transcription factors. Note that some families contain only a single known node. Figure 1d summarizes the hierarchy of physiological systems and organs into which experimental biosamples (tissues and cultured or primary cell lines) were classified. Consistent with terminology in use in the cellular signaling field^{1,53}, we refer to these individual gene products as nodes. Impacting the functions of nodes in all four categories are bioactive small molecules (BSMs), encompassing: 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 (Table 1). BSM-node mappings were retrieved from an existing pharmacology biocuration initiative, the IUPHAR Guide To Pharmacology⁵⁰, or annotated by SPP biocurators *de novo* with reference to a specific PubMed identifier (PMID).

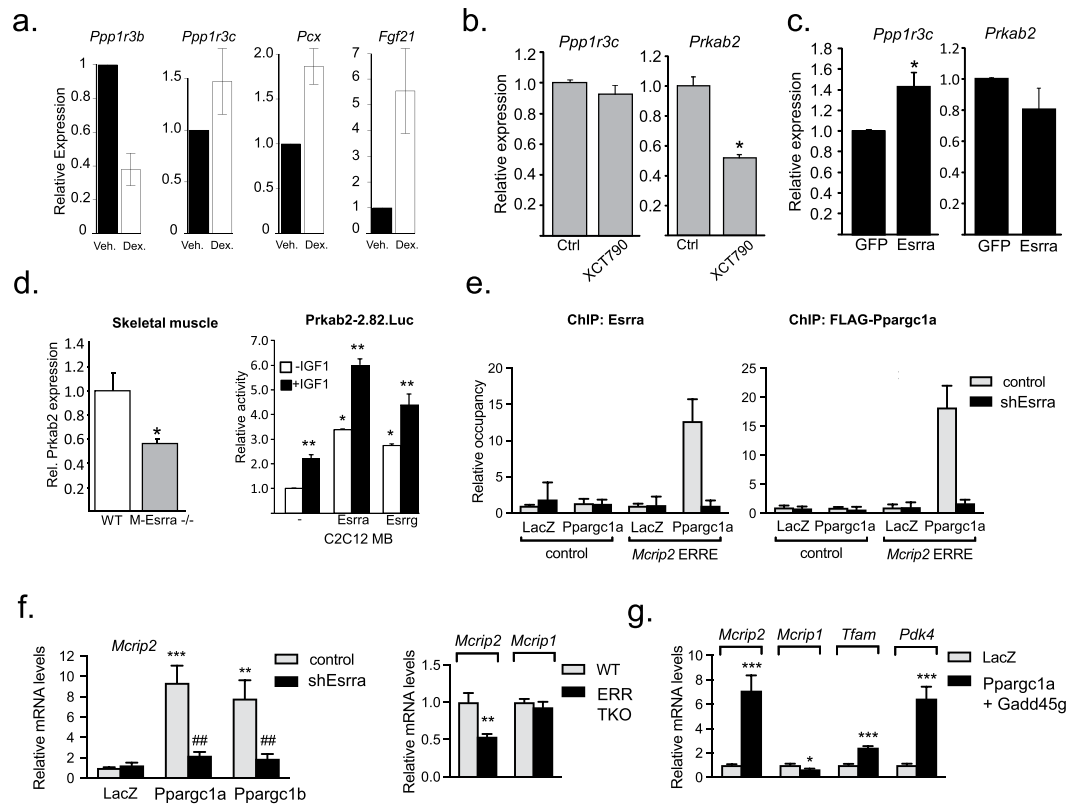


Fig. 7 Validation of consensome predictions of genomic targets for GR, ERR and IR. **(a)** *Ppp1r3c* is regulated by GR. Mouse Hepa-1-c hepatoma cells were treated with 250 nM dexamethasone (DEX) for 48 h, followed by qPCR of glycogenic genes including *Ppp1r3c*, *Ppp1r3b* and the established GR/NR3C1 targets pyruvate carboxylase (*Pcx*) and *Fgf21*. **(b)** Endogenous *Ppp1r3c* and *Prkab2* transcripts were measured by quantitative real-time PCR in C2C12 day 3 myotubes treated with vehicle or 5 μ M of the Esrra inverse agonist XCT790 (IC₅₀~0.5 mM) for 24 h. **(c)** Endogenous *Ppp1r3c* and *Prkab2* transcripts were measured by quantitative real-time PCR in C2C12 day 3 myotubes transduced with recombinant adenovirus expressing GFP or human Esrra/ERR α . Experimental transcript levels were normalized to 36B4 expression and results are expressed as the mean \pm S.E.M. Asterisks * indicate significant difference vehicle vs. treatment groups, ($p < 0.05$, $n = 3$). **(d)** Left panel. Expression of *Prkab2* transcript is reduced by 40% in Esrra-depleted skeletal muscle compared to wild-type tissue. Endogenous *Prkab2* expression was assayed by Q-PCR in vastus lateralis muscles of male wild-type (WT) or *Esrr1*^{-/-} mice (ERR α ^{-/-}). *Prkab2* transcript was normalized to 36B4 expression and results are expressed as the mean (\pm S.E.M). Asterisk * indicates significant difference between groups ($p < 0.05$, $n = 4$). Right panel. Activity of the *Prkab2*-2.82.Luc promoter-reporter in C2C12 myoblasts (MB) cotransfected with vector, ERR α /Esrra or ERR γ /Esrrg, as indicated. One day post-transfection MB were then cultured in 0.1% FBS overnight $-/+$ 10 nM IGF1 treatment for 24 hours. Data are reported as mean luciferase/renilla values normalized to control (\pm S.E.M.) for three trials. Asterisks indicate significant differences between transfection conditions (*) or IGF1 treatment (**), ($p \leq 0.05$, $n = 3$). **(e)** ChIP of Esrra (left panel) and Ppargc1a (right panel) at the *Mcrip2* ERRE. C2C12 myotubes were treated as described in Methods. Relative occupancy represents the amount of *Mcrip2* DNA (or of a control genomic region that has no ERR binding sites) that is immunoprecipitated by anti-Esrra or anti-Flag (detecting the Flag-tagged Ppargc1a in the different myotubes, relative to the DNA immunoprecipitated in LacZ/control shRNA cells (which has been set as 1 for each DNA region). Data are mean \pm SD ($n = 3$). **(f)** Left panel. *Mcrip2* is induced by Ppargc1 co-nodes in C2C12 myotubes in an Esrra-dependent manner. RNA (isolated 24 hrs after Ppargc1a/b expression) was analyzed by RT-qPCR. Data are normalized to 36B4, and expressed relative to levels in LacZ/shGFP cells. Right panel. Expression levels of *Mcrip2*, but not the related gene *Mcrip1*, are decreased in primary brown adipocytes lacking Esrra, Esrrb and Esrrg (ERR TKO), relative to ERR WT mice. mRNA levels were determined as in left panel. **(g)** Simulation of chronic adrenergic stimulation of primary brown adipocytes by overexpression of Ppargc1a and Gadd45g significantly increases expression of *Mcrip2* relative to mock-transfected adipocytes. Included as controls are the OXPHOS genes *Tfam* and *Pdk4*, encoding pyruvate dehydrogenase kinase isoform 4, a characterized ERR target and the third highest ranked target in the All nodes-Mm-adipose transcriptomic consensome signature. Differentiated adipocytes were infected with adenoviruses expressing Ppargc1a and Gadd45g and mRNA levels measured as described in the Methods section.

Dataset biocuration. 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 Atlas⁵. In order to expand this collection to encompass datasets involving perturbation of the full range of known signaling pathway nodes, we carried out a systematic survey of GEO and the research literature to identify an initial population of transcriptomic datasets representing a reasonable cross-section of the various classes of signaling pathway node referred to in Fig. 1. 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 at least two biological replicates to allow for calculation of 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 for failure to meet one or more of the above criteria.

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 investigators in their original 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 the previously described biosample controlled vocabulary. Model experiments representing a variety of physiological and metabolic processes (e.g. inflammation, adipogenesis, fasting-fed) were annotated where appropriate.

Gene differential expression values were calculated for each array and RNA-Seq experiment. **Array data.** Array data were processed as previously described⁵⁴. Briefly, expression data obtained from GEO are the investigator-provided summarized and normalized array feature expression intensities available in the “series matrix” or “processed” files, respectively. The full set of processed and normalized sample expression values provided by the investigator was extracted and processed in the current version of the statistical program R⁵⁵. To calculate differential gene expression for investigator-defined experimental contrasts, we used the linear modeling functions from the Bioconductor limma analysis package⁵⁶. Initially, a linear model was fitted to a group-means parameterization design matrix defining each experimental variable. Subsequently, we fitted a contrast matrix that recapitulated the sample contrasts of interest as defined in the study, producing fold-change and significance values for each array feature present on the array. P values obtained from Limma analysis were not corrected for multiple comparisons. In cases where a given gene was represented on an array by more than one probe-set, data from individual probe-sets were generated separately and fold-change values were not pooled across array features. **RNA-Seq data.** For RNA-Seq data, when aligned and annotated raw sequence count values were deposited to GEO, we used these along with R and the R limma and edgeR analysis packages to calculate differential expression values. Briefly, this involves the trimmed mean of M-values (TMM) normalization followed by voom transformation as detailed in the Limma user manual. Current BioConductor organism annotation libraries were used for annotation of investigator-provided gene identifiers. If RNA raw counts were not available in the GEO dataset record, we used BioJupies⁵⁷, which fetches the raw SRA FASTQ files and performs the alignment, annotation, and differential expression analysis. For both expression array and RNA-Seq data types, experiments were organized into datasets and assigned to newly minted digital object identifiers (DOIs), as previously described⁴.

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 resource⁷ 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 annotate ChIP-Atlas-processed SRA ChIP-Seq datasets relevant to mammalian signaling pathway nodes. Individual SRA experiments were first mapped to SRA Study Identifier (prefix SRP, DRP or ERP), which represents the SPP cistromic dataset unit. Individual SRA experiments (prefix SRX, DRX or ERX) were then mapped to the AGS of the immunoprecipitation (IP) node and any other genetically manipulated nodes (e.g. knockdown or knockout background), to any BSMs represented in the experimental design, and to the biosample in which the experiment was carried out. Experimental data and associated metadata were then loaded into the SPP Oracle 12c database.

Generation of consensomes. *Transcriptomic consensomes.* Transcriptomic ‘datasets’ are collections of data from different sources (i.e. different GEO datasets). Experiments, or contrasts in statistical terminology, are pairs of conditions (i.e. control and treatment) within data sets, each of which has multiple observations, that are used to generate nominal p-values and fold changes in expression for each gene (target) represented in the pair. These are all pre-computed and stored in the SPP Oracle database. Experiments are the unit of analysis, of which a single dataset can have one or more. differential expression values and associated significance measures were generated from appropriate experimental contrasts in GEO Series as previously described⁵⁴.

Large scale meta-analysis pipeline of publically archived transcriptomic datasets is complicated 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 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, prior to committing them to the consensome pipeline as described before.

Computation of target & experiment-specific nominal p-values & fold changes: Although RNA-Seq datasets are growing in number, expression arrays remain in use and the vast majority of expression profiling datasets archived in Gene Expression Omnibus are on array platforms. We first therefore set out to develop an algorithm that would establish consensus across array datasets. Although much less than 1% of genes in any particular array experiment are represented by more than 1 probeset, a few genes had 2–5 probesets and a very few had as many as 15 or 20. In such cases, we combined probeset-specific fold changes and probeset-specific p-values to generate gene level fold-changes and p-values. Briefly, we used the fold changes to convert the individual probeset-specific two-tailed nominal p-values into z-scores that capture the direction of the change:

$$Z_p = \text{qnorm} \left(\frac{1 + \text{sign}(\log_2(F_p)) - P_p}{2} \right)$$

where Z_p is the directional probeset-specific z-score, P_p is the two-tailed probeset-specific p-value, F_p is the probeset-specific fold change, $\text{qnorm}()$ is the standard normal inverse CDF, and $\text{sign}(x)$ is 1 when $x \geq 0$ and -1 when $x < 0$. Thus, when $F_p \geq 1$, this yields $Z_p = \text{qnorm}(1 - P_p/2)$ (range is $[0, \infty)$), and when $F_p < 1$, this yields the lower tail, $Z_p = \text{qnorm}(P_p/2)$ (range is $(-\infty, 0]$).

A summary gene-specific p-value was calculated as 2 times the upper tail of the standard normal cumulative distribution function assessed at the absolute value of the average of the probeset-specific Z 's:

$$Z = \frac{\sum_p Z_p}{n}$$

$$P = 2 * (1 - \text{pnorm}(|Z|))$$

where Z is the average of the probe-specific z-scores, P is the gene specific two-tailed p-value, n is the number of probesets for a gene, and pnorm is the standard normal cumulative distribution function.

The summary gene-specific experiment-specific fold change is calculated by exponentiating the predicted value of \log_2 fold change from a linear regression of probeset \log_2 fold changes regressed on probeset Z 's, evaluated at the average of the probeset Z 's:

$$\hat{F} = 2^{(\hat{a} + \hat{b} * Z)}$$

where \hat{F} is the predicted fold change at Z , \hat{a} is the intercept and \hat{b} is the slope of the linear model of $\log_2(F_p)$ modeled as a function of Z_p .

Combination of gene-specific p-values and fold changes across experiments: The calculation for a hypothetical target is shown in Supplementary Information Section 8. For each gene, g , in the consensome, we counted the number of experiments, E_g , where the gene has a nominal p-value of 0.05 or less out of N_g experiments where gene-specific data are present. The discovery rate (DR) was calculated as E_g/N_g . A consensus p-value, P_g (CPV) was calculated as the binomial probability of observing E_g or more successes out of N_g trials, when the true probability of success is 0.05. This provided an estimate of the degree to which the fraction of experiments with alterations exceed what might be expected by chance. The gene-specific consensus fold change, F_g , is geometric mean of the experiment-specific fold changes, expressed as $\max(F_{ge}, 1/F_{ge})$, for the gene of interest. 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 expression⁵⁸; cell cycle stage⁵⁹; biosample of study⁶⁰; 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, all fold changes were converted to $\max(F_{ge}, 1/F_{ge})$, such that both inductive and repressive experimental manipulations counted as 'altered', allowing us to generate a summary measure of magnitude of perturbation. Genes in the consensome analysis are ranked in ascending order by CPV, with average rank reported in the case of tied CPVs.

In summary, transcriptomic consensome analysis is predicated upon three assumptions:

- firstly, that borrowing statistical power by binning experiments according to their perturbation of a given signaling node is biologically valid;
- secondly, that omitting direction of differential expression from the analysis allows 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, more accurately reflects the strength of the regulatory relationship between a given node and its transcriptional targets.

Generation of cistromic consensomes. For calculation of ChIP-Seq consensomes, groups of experiments were designated 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. MACS2 peak calls from the ChIP-Atlas resource⁷ for all nodes in a defined SPP family were averaged and the targets ranked based upon this value.

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 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.

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.

SPP web application. The SPP knowledgebase is a gene-centric Java Enterprise Edition 6, web-based application around which other gene, mRNA, protein and BSM data from external databases such as NCBI are collected. After undergoing semiautomated processing and biocuration as described above, the data and annotations are stored in SPP's Oracle 12c database. RESTful web services exposing 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.

Bench validation and characterization experiments. *Validation and characterization of TPD52L1 in the ERs-Hs-MG consensome.* ER-Hs-MG 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 (0.5 µg) 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 $2^{-\Delta\Delta Ct}$ method⁶¹, and normalized to 36B4. All data shown is representative of at least three independent experiments. Primer sequences are shown in Supplementary Information Section 9.

Subcellular distribution: MCF-7 cells were kept in 5% CD-CS for 48 h prior treatment with 17BE2 10 nM for 24 h. A previously published immunofluorescence protocol was followed⁶². 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 4 °C 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 (0.25 µm steps covering 12 µm) 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 3×10^5 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 described⁶³. 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 $2^{-\Delta\Delta Ct}$ compared to the non-targeting siRNA control⁶⁴. 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 consensome. 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 1×10^6 and 2×10^6 , 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 µg 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'-TCAGACAGCTCTTTGGCTCA-3', reverse 5'-ACACCCCTGTTAGAAAACGTTAGAT-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 Δ Ct 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-Mm-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 48 h. 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 a 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 are shown in Supplementary Information Section 9.

Validation and characterization of Prkab2 in the ERRs-Mm-Metabolic consensome. 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)^{65,66}. 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 5 µM XCT790 (Sigma-Aldrich, St. Louis, MO), 0.1 µM DY40⁶⁷ 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-Esrrg, pSG5-HA-Esrra and pcDNA-myc/his-PGC-1α have been previously described⁶⁸. Transient transfection in C2C12 myocytes using the calcium phosphate method and the plasmid concentrations used have been described⁶⁹. Luciferase activity was assayed in MB 48 h post-transfection or in day 4 MT after changing confluent cells to 2% HS/DMEM. To assess IGF1 activation, MB were changed to SFM −/+ 10 nM 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 described⁶⁹. 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 described⁶⁹. 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 (\pm 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 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.

Validation and characterization of *Mcrip2* in the ERRs-Mm-Metabolic and PPARGC-Mm-Metabolic consensome. C2C12 myotube cultures: C2C12 myoblasts (ATCC) were plated at low density in 12-well tissue culture plates in complete DMEM (10% FBS), and switched to differentiation medium (DMEM supplemented with 2% horse serum) when cultures approached confluence. For *Esrra* knockdown experiments, myotubes were infected with adenoviruses expressing Ad-shERR α or vector control at m.o.i of 200 on day 4 of differentiation, and an additional dose of Ad-shERR α or vector control (at m.o.i of 100) together with adenoviruses expressing LacZ (control), *Ppargc1a* or *Ppargc1b* (at m.o.i of 50) on day 6 of differentiation. 24 h later, RNA or DNA was harvested for RT-qPCR gene expression or ChIP analyses, respectively. For expression of the constitutively active *Esrra* (VP16-ERR α), *Esrrb* or *Esrrg*, day 6 myotubes were infected with MOI 50 of adenoviruses expressing LacZ (control), VP16-ERR α , ERR β or ERR γ ^{32,70}. RNA was harvested 24 h later.

Primary brown adipocyte cultures: Pre-adipocytes were isolated from the BAT depot of mice with floxed ERR alleles, as previously described⁷¹ and cultured in DMEM supplemented with 20 mM HEPES and 20% FBS prior to differentiation. To induce recombination of floxed ERR loci, pre-adipocytes at 70% confluency were incubated for 16 h with GFP- (control) or CRE- expressing lentiviruses in media containing 4 μ g/ml polybrene⁷¹. Upon confluency (day 0), cells were switched to DMEM supplemented with 10% FBS, 20 nM insulin, 1 nM triiodothyronine, 0.5 mM IBMX, 2 μ g/ml DEX and 0.25 mM indomethacin to differentiate. On day 2 of differentiation, cells were switched to DMEM supplemented with 10% FBS, 20 nM insulin, and 1 nM triiodothyronine. For overexpression assays, mature adipocytes were infected on day 5 of differentiation with adenoviruses expressing *Ppargc1a* or *Gadd45g* at an m.o.i of 20. RNA was harvested 24 h later.

Q-PCR: Quantitative RT-PCR was performed using the following gene-specific primers: *Rplp0* (36B4), CTGTGCCAGCTCAGAACACTG and TGATCAGCCCGAAGAGAAG; *Mcrip2*, GCCTGTGCAGTATGTGGAGA and GGGTCCACTATGGCAACATT; *Mcrip1*, AAGAGAATGTGCGCTTCATTTA and CTAGGCACCGCTCACCAC; *Pdk4*, GTTCCTCACACCTTACCAC and CCTCCTCGGTCAGAAATCTTG; *Tfam*, CAAAGGATGATTTCGGCTCAG and AAGCTGAATATATGCCTGCTTTTC. Relative mRNA expression was normalized using *Rplp0* (36B4) as reference gene.

ChIP: C2C12 myotubes were crosslinked for 10 min at 37° in 1% formaldehyde in PBS. After quenching, sonication to ~500 bp fragments, and pre-clearing by treatment with protein A/G sepharose, soluble chromatin was immunoprecipitated with antibodies against *Esrra* or FLAG. Immunoprecipitated DNA was quantified by real-time PCR, using primers for the *Mcrip2* ERRE (TGAGTACTTGCGGTCCTTGA and ACCTTGGAGAAGGTTGATGG), or primers for a region distal to the *Esrra* promoter that lacks ERREs (negative control; primers described in⁷²). Data shown are the mean and standard deviation of three experimental replicates.

Data Availability

SPP is freely accessible at <https://www.signalingpathways.org>. Programmatic access to all underlying data points and their associated metadata are supported by a RESTful API at <https://www.signalingpathways.org/docs/>. All SPP datasets are biocurated versions of publically archived datasets, are formatted according to the recommendations of the FORCE11 Joint Declaration on Data Citation Principles⁷³, and are made available under a Creative Commons CC 3.0 BY license. The original datasets are available and linked to from the corresponding SPP datasets using the original repository accession identifiers. These identifiers are for transcriptomic datasets, the Gene Expression Omnibus (GEO) Series (GSE); and for cistromic/ChIP-Seq datasets, the NCBI Sequence Read Archive (SRA) study identifier (SRP).

Code availability

The full SPP source code is available in the SPP GitHub account under a Creative Commons CC 3.0 BY license at <https://github.com/signaling-pathways-project/ominer/>.

Received: 22 February 2019; Accepted: 11 September 2019;

Published online: 31 October 2019

References

1. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* **7**, 85–96, <https://doi.org/10.1038/nrm1837> (2006).
2. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* **44**, D481–487, <https://doi.org/10.1093/nar/gkv1351> (2016).
3. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18> (2016).
4. Darlington, Y. F. *et al.* Improving the discoverability, accessibility, and citability of omics datasets: a case report. *J Am Med Inform Assoc* **24**, 388–393, <https://doi.org/10.1093/jamia/ocw096> (2017).

5. Becnel, L. B. *et al.* Discovering relationships between nuclear receptor signaling pathways, genes, and tissues in Transcriptomine. *Sci Signal* **10**, <https://doi.org/10.1126/scisignal.aah6275> (2017).
6. McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474 (2002).
7. Oki, S. *et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* **19**, <https://doi.org/10.15252/embr.201846255> (2018).
8. Planque, R., Hulshof, J., Teusink, B., Hendriks, J. C. & Bruggeman, F. J. Maintaining maximal metabolic flux by gene expression control. *PLoS Comput Biol* **14**, e1006412, <https://doi.org/10.1371/journal.pcbi.1006412> (2018).
9. Desvergne, B., Michalik, L. & Wahli, W. Transcriptional regulation of metabolism. *Physiol Rev* **86**, 465–514, <https://doi.org/10.1152/physrev.00025.2005> (2006).
10. Amberger, J. S. & Hamosh, A. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinformatics* **58**, 1.2.1–1.2.12, <https://doi.org/10.1002/cpbi.27> (2017).
11. Zhao, M. & Qu, H. Human liver rate-limiting enzymes influence metabolic flux via branch points and inhibitors. *BMC Genomics* **10**(Suppl 3), S31, <https://doi.org/10.1186/1471-2164-10-S3-S31> (2009).
12. Shehata, M., Weidenhofer, J., Thamocharampillai, K., Hardy, J. R. & Byrne, J. A. Tumor protein D52 overexpression and gene amplification in cancers from a mosaic of microarrays. *Crit Rev Oncol* **14**, 33–55 (2008).
13. Toyoshima, F. & Nishida, E. Integrin-mediated adhesion orients the spindle parallel to the substratum in an EB1- and myosin X-dependent manner. *Embo j* **26**, 1487–1498, <https://doi.org/10.1038/sj.emboj.7601599> (2007).
14. Hishikawa, D. *et al.* Discovery of a lysophospholipid acyltransferase family essential for membrane asymmetry and diversity. *Proc Natl Acad Sci USA* **105**, 2830–2835, <https://doi.org/10.1073/pnas.0712245105> (2008).
15. Han, W. *et al.* Reactivation of androgen receptor-regulated lipid biosynthesis drives the progression of castration-resistant prostate cancer. *Oncogene* **37**, 710–721, <https://doi.org/10.1038/ncr.2017.385> (2018).
16. Tsai, H. C., Boucher, D. L., Martinez, A., Tepper, C. G. & Kung, H. J. Modeling truncated AR expression in a natural androgen responsive environment and identification of RHOB as a direct transcriptional target. *PLoS One* **7**, e49887, <https://doi.org/10.1371/journal.pone.0049887> (2012).
17. Cunha, G. R. *et al.* The endocrinology and developmental biology of the prostate. *Endocr Rev* **8**, 338–362, <https://doi.org/10.1210/edrv-8-3-338> (1987).
18. Tabe, S. *et al.* Lysophosphatidylcholine acyltransferase 4 is involved in chondrogenic differentiation of ATDC5 cells. *Sci Rep* **7**, 16701, <https://doi.org/10.1038/s41598-017-16902-4> (2017).
19. Koelling, S. & Miosge, N. Sex differences of chondrogenic progenitor cells in late stages of osteoarthritis. *Arthritis Rheum* **62**, 1077–1087, <https://doi.org/10.1002/art.27311> (2010).
20. Fong, N. M. *et al.* Identification of binding sites on protein targeting to glycogen for enzymes of glycogen metabolism. *J Biol Chem* **275**, 35034–35039, <https://doi.org/10.1074/jbc.M005541200> (2000).
21. Jeon, S. M. Regulation and function of AMPK in physiology and diseases. *Exp Mol Med* **48**, e245, <https://doi.org/10.1038/emm.2016.81> (2016).
22. Vanstapel, F., Dopere, F. & Stalmans, W. The role of glycogen synthase phosphatase in the glucocorticoid-induced deposition of glycogen in foetal rat liver. *Biochem J* **192**, 607–612 (1980).
23. Huss, J. M., Garbacz, W. G. & Xie, W. Constitutive activities of estrogen-related receptors: Transcriptional regulation of metabolism by the ERR pathways in health and disease. *Biochim Biophys Acta* **1852**, 1912–1927, <https://doi.org/10.1016/j.bbadis.2015.06.016> (2015).
24. Muhic, M., Vardjan, N., Chowdhury, H. H., Zorec, R. & Kreft, M. Insulin and Insulin-like Growth Factor 1 (IGF-1) Modulate Cytoplasmic Glucose and Glycogen Levels but Not Glucose Transport across the Membrane in Astrocytes. *J Biol Chem* **290**, 11167–11176, <https://doi.org/10.1074/jbc.M114.629063> (2015).
25. Printen, J. A., Brady, M. J. & Saltiel, A. R. PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. *Science* **275**, 1475–1478 (1997).
26. O'Doherty, R. M. *et al.* Activation of direct and indirect pathways of glycogen synthesis by hepatic overexpression of protein targeting to glycogen. *J Clin Invest* **105**, 479–488, <https://doi.org/10.1172/jci8673> (2000).
27. Dasgupta, B. *et al.* The AMPK beta2 subunit is required for energy homeostasis during metabolic stress. *Mol Cell Biol* **32**, 2837–2848, <https://doi.org/10.1128/MCB.05853-11> (2012).
28. Kim, J., Yang, G., Kim, Y., Kim, J. & Ha, J. AMPK activators: mechanisms of action and physiological activities. *Exp Mol Med* **48**, e224, <https://doi.org/10.1038/emm.2016.16> (2016).
29. Aghanoori, M. R. *et al.* Insulin-like growth factor-1 activates AMPK to augment mitochondrial function and correct neuronal metabolism in sensory neurons in type 1 diabetes. *Mol Metab* **20**, 149–165, <https://doi.org/10.1016/j.molmet.2018.11.008> (2019).
30. Altshuler-Keylin, S. & Kajimura, S. Mitochondrial homeostasis in adipose tissue remodeling. **10**, <https://doi.org/10.1126/scisignal.aai9248> (2017).
31. Cannon, B. & Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiol Rev* **84**, 277–359, <https://doi.org/10.1152/physrev.00015.2003> (2004).
32. Gantner, M. L., Hazen, B. C., Conkright, J. & Kralli, A. GADD45gamma regulates the thermogenic capacity of brown adipose tissue. *Proc Natl Acad Sci USA* **111**, 11870–11875, <https://doi.org/10.1073/pnas.1406638111> (2014).
33. Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab* **1**, 361–370, <https://doi.org/10.1016/j.cmet.2005.05.004> (2005).
34. Bish, R. *et al.* Comprehensive Protein Interactome Analysis of a Key RNA Helicase: Detection of Novel Stress Granule Proteins. *Biomolecules* **5**, 1441–1466, <https://doi.org/10.3390/biom5031441> (2015).
35. Hernandez-de-Diego, R. *et al.* PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res* **46**, W503–w509, <https://doi.org/10.1093/nar/gky466> (2018).
36. Perez-Riverol, Y. *et al.* Discovering and linking public omics data sets using the Omics Discovery Index. *Nat Biotechnol* **35**, 406–409, <https://doi.org/10.1038/nbt.3790> (2017).
37. McArdle, S. *et al.* PRESTO, a new tool for integrating large-scale -omics data and discovering disease-specific signatures. *BioRxiv*, <https://doi.org/10.1101/302604> (2018).
38. Parikh, J. R., Klinger, B., Xia, Y., Marto, J. A. & Bluthgen, N. Discovering causal signaling pathways through gene-expression patterns. *Nucleic Acids Res* **38**, W109–117, <https://doi.org/10.1093/nar/gkq424> (2010).
39. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452, <https://doi.org/10.1016/j.cell.2017.10.049> (2017).
40. 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 Res* **46**, D267–d275, <https://doi.org/10.1093/nar/gkx1092> (2018).
41. Wang, Z. *et al.* Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat Commun* **7**, 12846, <https://doi.org/10.1038/ncomms12846> (2016).
42. Cantini, L. *et al.* Classification of gene signatures for their information value and functional redundancy. *NPJ Syst Biol Appl* **4**, 2, <https://doi.org/10.1038/s41540-017-0038-8> (2018).

43. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* **9**, 20, <https://doi.org/10.1038/s41467-017-02391-6> (2018).
44. Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res* **45**, D972–d978, <https://doi.org/10.1093/nar/gkw838> (2017).
45. Yoo, M. *et al.* DSigDB: drug signatures database for gene set analysis. *Bioinformatics* **31**, 3069–3071, <https://doi.org/10.1093/bioinformatics/btv313> (2015).
46. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425, <https://doi.org/10.1016/j.cels.2015.12.004> (2015).
47. 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 Syst Biol* **5**, 138, <https://doi.org/10.1186/1752-0509-5-138> (2011).
48. Stanislawska-Sachadyn, A., Sachadyn, P. & Limon, J. Transcriptomic Effects of Estrogen Starvation and Induction in the MCF7 Cells. The Meta-analysis of Microarray Results. *Curr Pharm Biotechnol* **17**, 161–172 (2015).
49. Thomas, S. M. & Brugge, J. S. Cellular functions regulated by Src family kinases. *Annu Rev Cell Dev Biol* **13**, 513–609, <https://doi.org/10.1146/annurev.cellbio.13.1.513> (1997).
50. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res* **44**, D1054–1068, <https://doi.org/10.1093/nar/gkv1037> (2016).
51. Webb, E. C. *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes.* (Academic Press, 1992).
52. Wingender, E., Schoeps, T. & Donitz, J. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res* **41**, D165–170, <https://doi.org/10.1093/nar/gks1123> (2013).
53. Minard, A. Y. *et al.* mTORC1 Is a Major Regulatory Node in the FGF21 Signaling Network in Adipocytes. *Cell Rep* **17**, 29–36, <https://doi.org/10.1016/j.celrep.2016.08.086> (2016).
54. Ochsner, S. A. *et al.* Transcriptome, a web resource for nuclear receptor signaling transcriptomes. *Physiol Genomics* **44**, 853–863, <https://doi.org/10.1152/physiolgenomics.00033.2012> (2012).
55. Ihaka, R. & Gentleman, R. R. R: a language for data analysis and graphics. *J Computational Graph Stat* **5**, 299–314 (1996).
56. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article 3, <https://doi.org/10.2202/1544-6115.1027> (2004).
57. Torre, D., Lachmann, A. & Ma'ayan, A. BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud. *Cell Syst* **7**, 556–561 e553, <https://doi.org/10.1016/j.cels.2018.10.007> (2018).
58. Foulkes, N. S. & Sassone-Corsi, P. More is better: activators and repressors from the same gene. *Cell* **68**, 411–414 (1992).
59. Bertoli, C., Skotheim, J. M. & de Bruin, R. A. Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol* **14**, 518–528, <https://doi.org/10.1038/nrm3629> (2013).
60. Okabe, Y. & Medzhitov, R. Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell* **157**, 832–844, <https://doi.org/10.1016/j.cell.2014.04.016> (2014).
61. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-(Delta Delta C(T))} Method. *Methods* **25**, 402–408, <https://doi.org/10.1006/meth.2001.1262> (2001).
62. Stossi, F. *et al.* High throughput microscopy identifies bisphenol AP, a bisphenol A analog, as a novel AR down-regulator. *Oncotarget* **7**, 16962–16974, <https://doi.org/10.18632/oncotarget.7655> (2016).
63. Foulds, C. E. *et al.* Research resource: expression profiling reveals unexpected targets and functions of the human steroid receptor RNA activator (SRA) gene. *Mol Endocrinol* **24**, 1090–1105, <https://doi.org/10.1210/me.2009-0427> (2010).
64. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101–1108 (2008).
65. Huss, J. M. *et al.* The nuclear receptor ERRalpha is required for the bioenergetic and functional adaptation to cardiac pressure overload. *Cell Metab* **6**, 25–37, <https://doi.org/10.1016/j.cmet.2007.06.005> (2007).
66. Luo, J. *et al.* Reduced fat mass in mice lacking orphan nuclear receptor estrogen-related receptor alpha. *Mol Cell Biol* **23**, 7947–7956 (2003).
67. Yu, D. D., Huss, J. M., Li, H. & Forman, B. M. Identification of novel inverse agonists of estrogen-related receptors ERRgamma and ERRbeta. *Bioorg Med Chem* **25**, 1585–1599, <https://doi.org/10.1016/j.bmc.2017.01.019> (2017).
68. Huss, J. M., Kopp, R. P. & Kelly, D. P. 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. *J Biol Chem* **277**, 40265–40274, <https://doi.org/10.1074/jbc.M206324200> (2002).
69. Murray, J. & Huss, J. M. Estrogen-related receptor alpha regulates skeletal myocyte differentiation via modulation of the ERK MAP kinase pathway. *Am J Physiol Cell Physiol* **301**, C630–645, <https://doi.org/10.1152/ajpcell.00033.2011> (2011).
70. Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T. & Kralli, A. The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor estrogen-related receptor alpha (ERRalpha). *J Biol Chem* **278**, 9013–9018, <https://doi.org/10.1074/jbc.M212923200> (2003).
71. Gantner, M. L., Hazen, B. C., Eury, E., Brown, E. L. & Kralli, A. Complementary Roles of Estrogen-Related Receptors in Brown Adipocyte Thermogenic Function. *Endocrinology* **157**, 4770–4781, <https://doi.org/10.1210/en.2016-1767> (2016).
72. Villena, J. A. *et al.* Orphan nuclear receptor estrogen-related receptor alpha is essential for adaptive thermogenesis. *Proc Natl Acad Sci USA* **104**, 1418–1423, <https://doi.org/10.1073/pnas.0607696104> (2007).
73. Cousijn, H. *et al.* A data citation roadmap for scientific publishers. *Sci Data* **5**, 180259, <https://doi.org/10.1038/sdata.2018.259> (2018).

Acknowledgements

This work was supported by the National Institute of Diabetes, Digestive and Kidney Diseases (DK097771, DK097748, DK48807, DK107535, DK56338, DK095686 and DK105126), the National Institute of Child Health and Development (DK097748), the National Cancer Institute (CA125123) 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, and by a grant from the American Thyroid Association. Microscopic analysis was supported by the Integrated Microscopy Core at Baylor College of Medicine with funding from the Cancer Prevention Research Institute of Texas (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 whom SPP would not be possible.

Author contributions

SPP knowledgebase concept and design: N.M. and S.O. Consensus algorithm design: S.H., N.M. and S.O. Data processing: S.O., Z.W. and A.M. Biocuration: N.M., S.A.O., D.A. and K.M. Software development: L.B., N.M., A.M., W.D. and W.K. Oracle database and system administration: S.Q. and M.D. Bench validation: K.A., R.H., Y.C., A.H., M.G., D.B., J.H., F.S., C.F., A.K. and D.M. Manuscript drafting: N.M., D.B., J.H., C.F., A.K. and D.M.

Competing interests

Charles Foulds has equity in Coactigon, Inc.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41597-019-0193-4>.

Correspondence and requests for materials should be addressed to N.J.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019