Supplemental Materials

Includes:
Supplemental Figures
Supplemental Materials and Methods
Supplemental References

Identification and characterization of the Mediator kinase-dependent myometrial stem cell phosphoproteome

Lindsey Barron, Subash Khadka, Robert Schenken, Long He, John Blenis, Julian Blagg, Shin-Fu Chen, Kuang-Lei Tsai and Thomas G. Boyer

Supplemental Figure 1

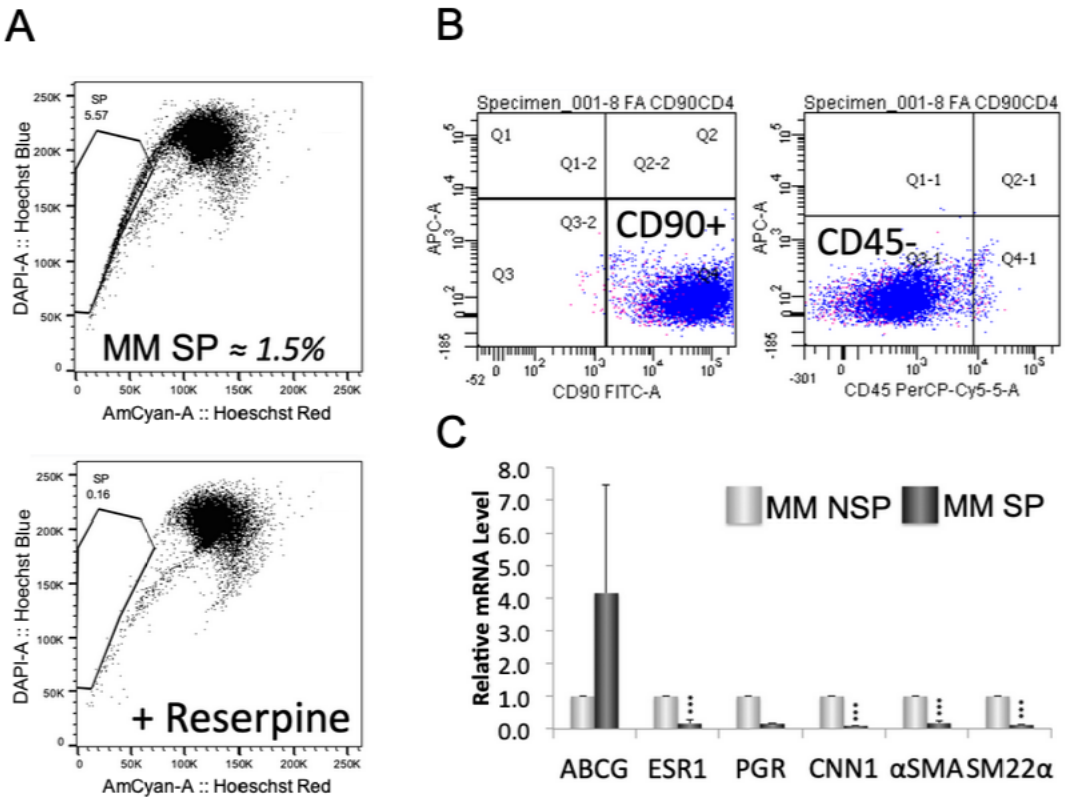

Figure S1. Isolation and characterization of side population (SP) cells from patient-derived myometrium. (**A, *top***) Distribution of SP (framed) and non-SP cells within all Hoechst-stained live cells isolated from patient-derived myometrial tissue. (**A, bottom)** Addition of Reserpine (ABC transport inhibitor; 50μM) vanquishes SP fractions. (**B**) Expression of bone marrow mesenchymal (CD90) and hematopoietic (CD45) stem cell markers in myometrium-derived SP cells. (**C**) RT-qPCR was used to quantify mRNA expression levels of stem cell (ABCG2), nuclear hormone receptor [estrogen receptor α (ESR1), progesterone receptor (PGR)], and smooth muscle [calponin 1 (CNN1), α smooth muscle actin (αSMA; aka ATCA2), transgelin (SM22α)] markers in myometrial non-SP (MM NSP) and SP (MM SP) cells. mRNA expression levels were normalized to that of GAPDH, and expressed relative to their levels in MM NSP cells. Data are mean ± SEM of 3 experiments in triplicate. Asterisks denote significant differences (Student's *t*-test: ***$p<0.001$).

Supplemental Figure 2

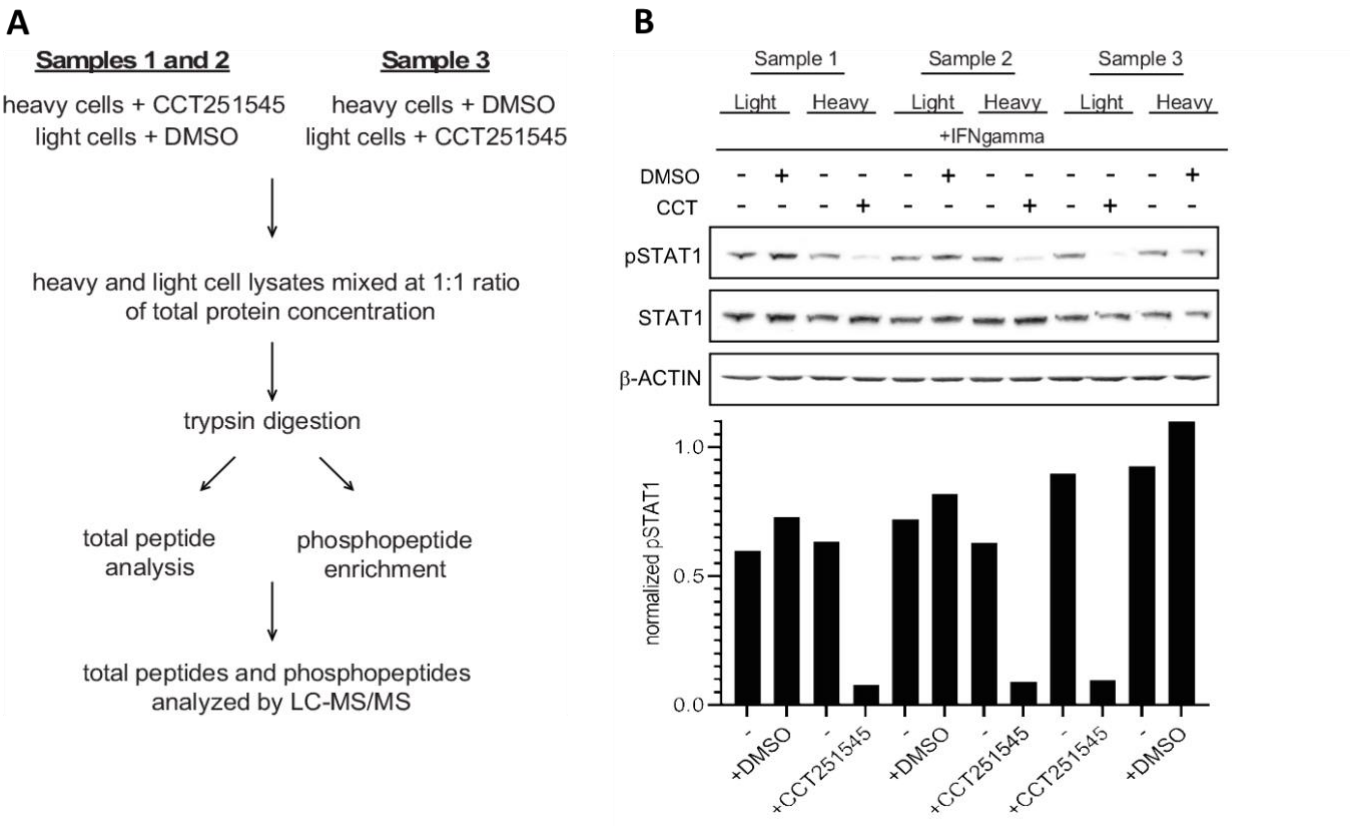

Figure S2. Study design and validation for SILAC-based phosphoproteomics coupled with chemical inhibition of Mediator kinase activity. (**A**) Schematic summary of SILAC phosphoproteomics workflow. (**B**) Validation of Mediator kinase inhibition by CCT251545 in MM SP cells. Replicate cultures of metabolically labeled MM SP cells used for SILAC analysis as in (**A**) were processed in parallel for CDK8/19 inhibition by monitoring the phosphorylation level STAT1 S727, an established Mediator kinase substrate. Metabolically labeled cells were treated with DMSO or CCT251545 (100 nM) as indicated along with γ-interferon (10ng/ml) for 45 minutes prior to cell harvest and processing of whole cell lysates for immunoblot analysis using the indicated antibodies specific for phosphorylated STAT1 S727 (pSTAT1), bulk STAT1, and β-ACTIN, the latter of which served as an internal loading control. The level of phosphorylated STAT1 under each condition was determined by quantification of immunoblot signals and normalization of the pSTAT1 signal to that of both bulk STAT1 and β-ACTIN. Normalized pSTAT1 levels under each condition are plotted in the bar graph below the immunoblot.

Supplemental Figure 3

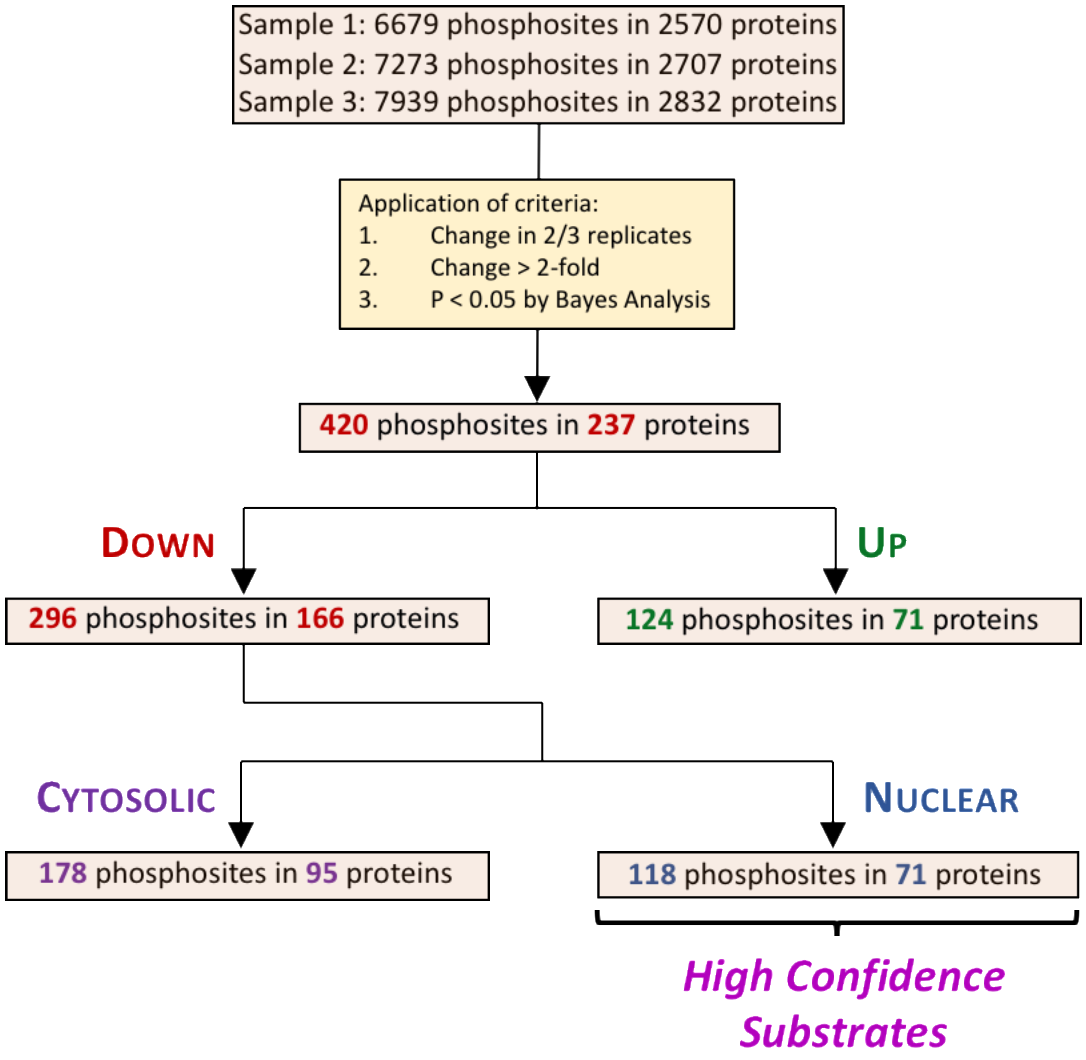

Figure S3. Stratification strategy for identification of "high-confidence" Mediator kinase substrates. Mediator kinase-dependent phosphosites identified by SILAC and quantitative phosphoproteomics workflow were deemed high-confidence by application of specific criteria (boxed) related to the magnitude and significance of inhibitor-dependent changes in protein phosphorylation levels across multiple experiments. Phosphosites within nuclear proteins exhibiting an inhibitor-dependent decrease in phosphorylation (118 phosphosites in 71 proteins) were deemed "high-confidence" Mediator kinase substrates.

Supplemental Figure 4

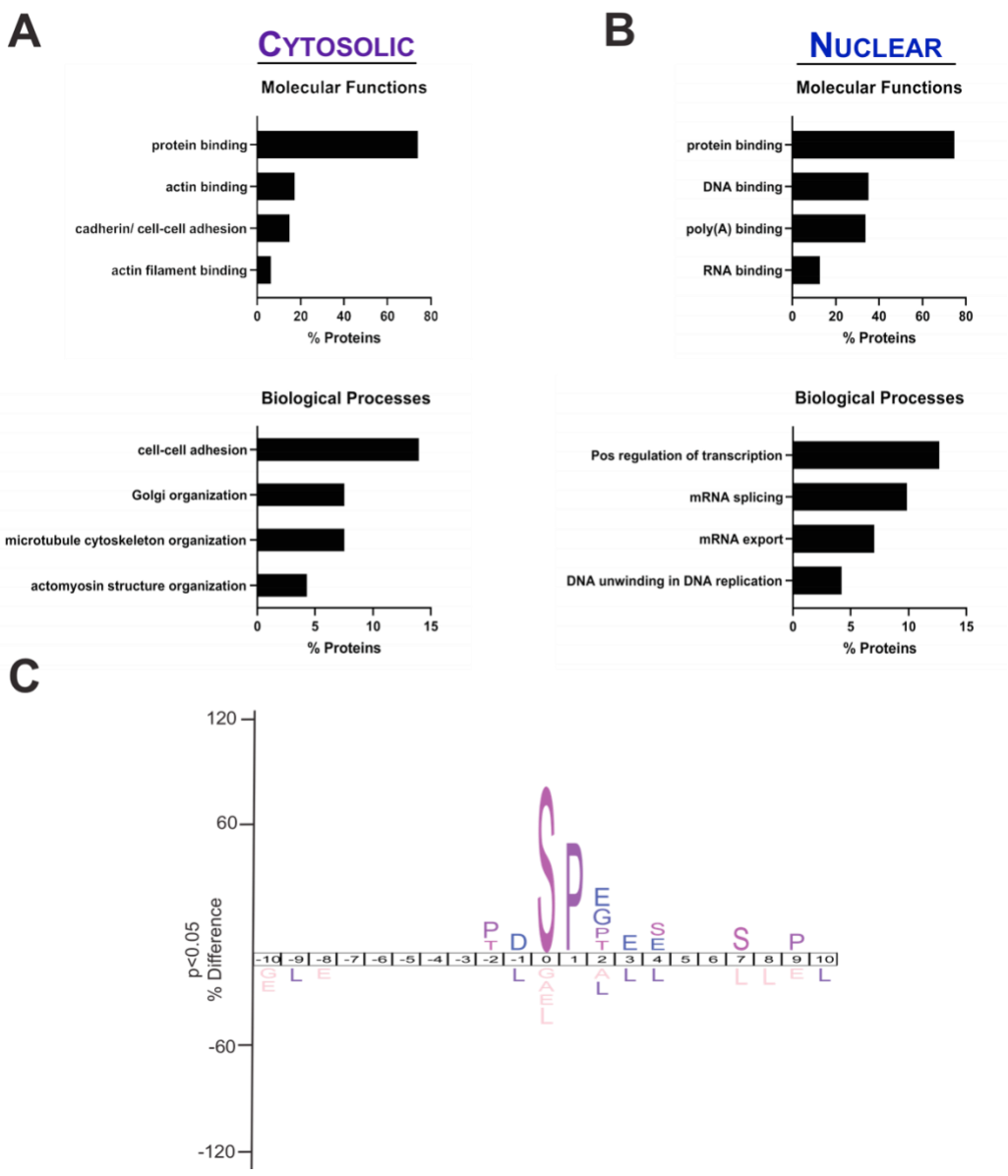

Figure S4. Gene ontological classification and sequence motif analysis for high-confidence Mediator kinase targets. The DAVID bioinformatics tool was used to identify enriched GO terms related to Molecular Functions and Biological Processes for Mediator kinase-dependent phosphoproteins localized predominantly to the cytosol (**A**) or nucleus (**B**). Shown are enriched terms with *p* values and FDRs of $\leq$ 0.05). (**C**) iceLogo was used to identify the consensus motif for high-confidence Mediator kinase substrates (listed in Table 1). Sequence logo shows over-and under-represented amino acids surrounding the mapped Mediator kinase phosphosites. Percent difference over background of enriched amino acids with a p<0.05 are shown.

Supplemental Figure 5

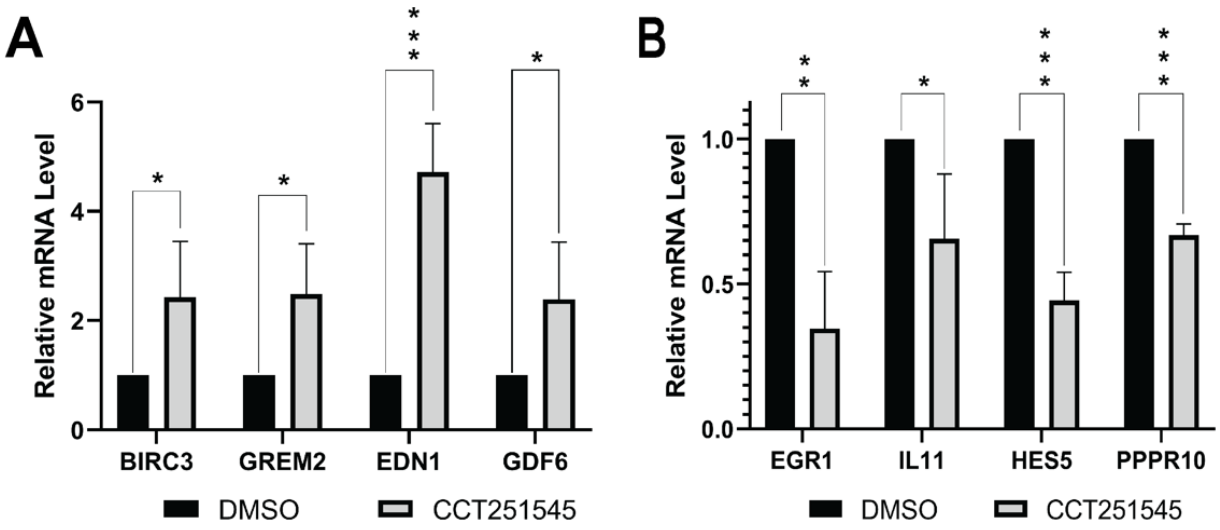

Figure S5. Validation of RNA-seq data. RT-qPCR was used to validate the indicated subset of genes identified by RNA-seq to be upregulated (**A**) or downregulated (**B**) following 3 hour treatment of MM SP cells with Mediator kinase inhibitor CCT251545 (100 nM). The mRNA expression level for each gene was normalized to that of GAPDH, and expressed relative to its normalized level in control (DMSO-treated) MM SP cells. Data are mean ± SEM of 3 experiments performed in triplicate. Asterisks denote significant differences (Student's *t*-test: *p<0.05, **p<0.01, ***p<0.001).

Supplemental Figure 6

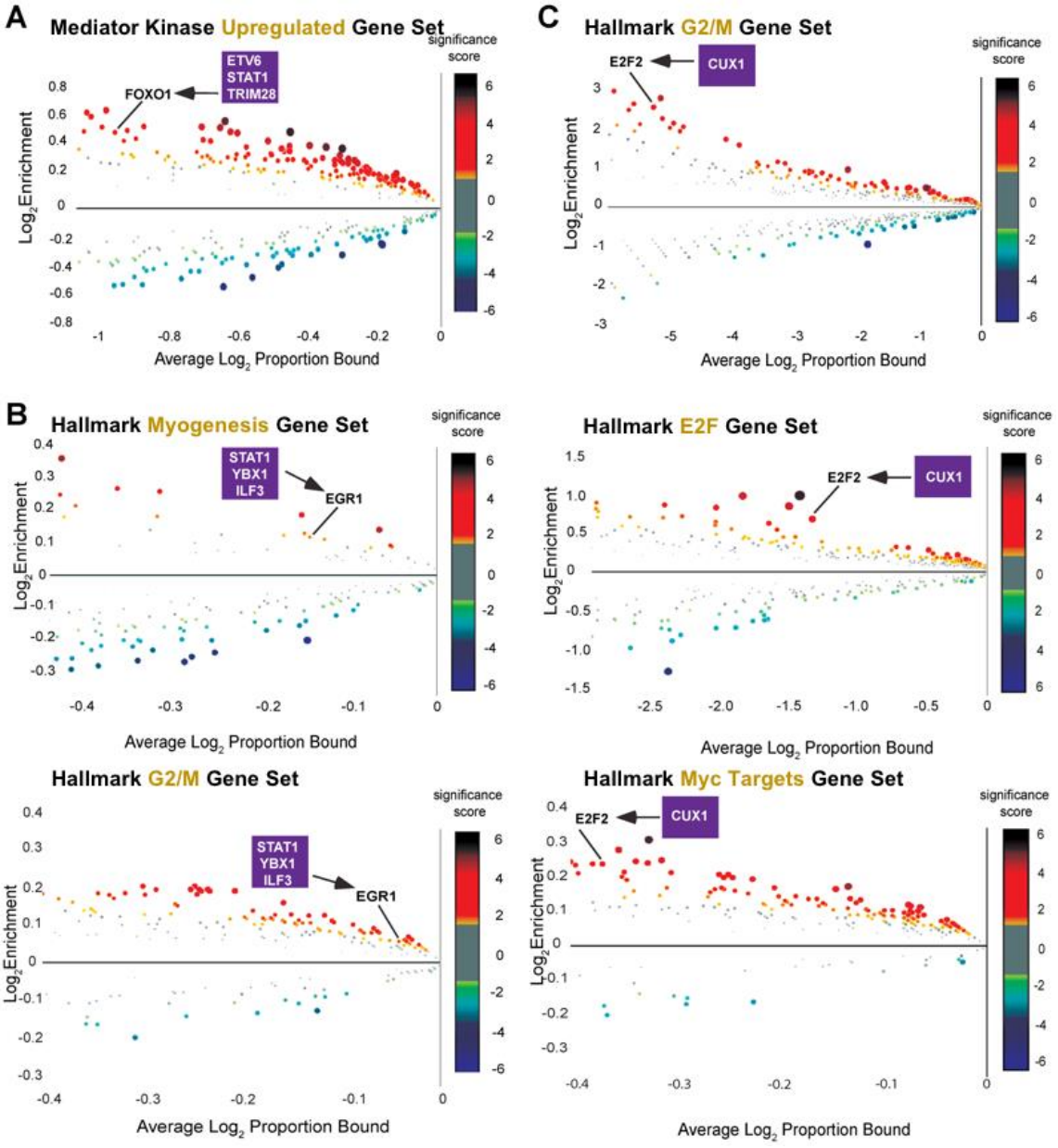

Figure S6. Mediator kinase substrates are co-regulators of Mediator kinase regulated genes in MM SP cells. Ciiider enrichment plots for Mediator kinase regulated genes. Ciiider is a computational transcription factor binding site (TFBS) prediction toolkit. Plots show the enrichment [proportional ratio of promoter regions bound by transcription factors (indicated by dots)] and average log proportion bound. Dot size and color reflect the Log_2 *p* value (significance score). Significance scores greater than or less than zero indicate overrepresentation or underrepresentation of TFBS, respectively, in queried gene sets. (**A**) Binding sites for FOXO1, a Mediator kinase regulated gene that physically and functionally interacts with Mediator kinase substrates ETV6, STAT1, and TRIM28, are enriched in Mediator kinase upregulated gene set

(n=123, from Fig. 3A and Table S4). **(B**) Binding sites for EGR1, a Mediator kinase regulated gene that physically and functionally interacts with Mediator kinase substrates STAT1, YBX1, and ILF3, are enriched in Mediator kinase regulated Hallmark Myogenesis (top) and G2/M Checkpoint (bottom) gene sets. (**C**) Binding sites for E2F2, a Mediator kinase regulated gene that physically and functionally interacts with Mediator kinase substrate CUX1, are enriched in Mediator kinase regulated Hallmark G2/M (top), E2F (middle), and MYC Target (bottom) gene sets.

Supplemental Figure 7

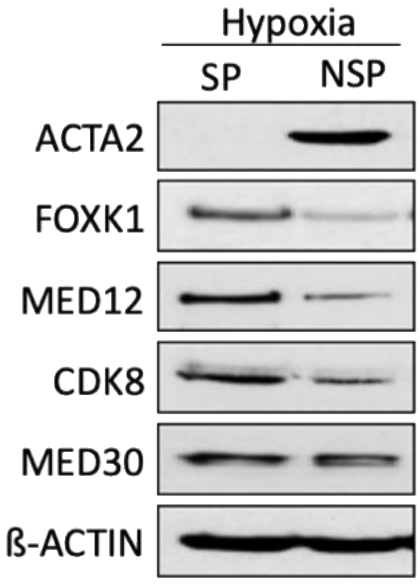

Figure S7. Mediator kinase module components and FOXK1 are coordinately enriched in MM SP cells compared to MM non-SP cells. Whole cell lysates of paired SP and non-SP cells from patient-derived myometrium (patient 1 in Figure 4) were cultured under hypoxia and processed by SDS-PAGE and immunoblot analysis using antibodies specific for: Mediator kinase substrate FOX1, Mediator kinase module subunits (MED12, CDK8), Mediator core subunit (MED30), α-smooth muscle actin (ACTA2), and β-ACTIN (loading control).

Supplemental Materials and Methods

Isolation of Myometrium Side Population Cells

Myometrium was minced manually into small pieces (1–2 mm^3) and incubated in maintenance medium (M199 medium (Invitrogen) containing 10% fetal bovine serum (FBS) (Gibco), antibiotic-antimycotic solution (Thermo-Fisher Scientific), and 10 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) buffer solution (Invitrogen). Samples were then digested at 37°C overnight with agitation by enzymatic means (2 mg/mL type II collagenase (Labclinics) and 1 mg/mL Dnase I (Sigma-Aldrich) to obtain single-cell suspensions. The cell suspensions were filtered through 70μm cell strainer (Greiner) and treated with erythrocyte lysis solution (ammonium chloride, potassium bicarbonate, tetrasodium ethylene-diamine tetra-acetic acid). Cell viability was performed with a cytometric analysis using propidium iodide (PI) 5 μg/mL (Sigma-Aldrich).

The cells were suspended at a concentration of 1×10^7 cells/mL in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 2% FBS (Gibco) and 10 mM HEPES (Sigma-Aldrich). Cells were labeled with 5 mg/mL of Hoechst 33342 dye (Ho-33342; Sigma-Aldrich) in a water bath at 37°C for 90–120min. A parallel aliquot (negative control) was stained with Hoechst dye in the presence of 50 μM reserpine (Sigma-Aldrich). After Hoechst incubation, cells were centrifuged at 4°C and resuspended in cold Hanks' balanced salt solution (HBSS) (Gibco) supplemented with 2% FBS and 10 mM HEPES and 1 μg/mL PI (Sigma-Aldrich).

Sample Preparation for LC-MS/MS

Proteins were extracted from isotopically labeled cells using protein lysis buffer (8M urea, 1% protease inhibitor, 1% phosphatase inhibitor) and sonication. The protein lysates were centrifuged at 12,000g at 4°C for 10 min. The supernatant was transferred to a new centrifuge tube and the protein concentration was determined by a BCA colorimetric assay. Heavy and light cell lysates for each respective sample were then mixed at a 1:1 ratio of protein concentration. In-solution digestion was performed on the sample by the addition of 50 mM ammonium bicarbonate and free trypsin to the protein solution at a ratio of 1:50. The digestion was then

incubated at 37°C overnight. At this point an aliquot of digested peptides was saved for total-protein profiling by LC-MS/MS. The remaining digested peptides were subjected to phospho-peptide enrichment.

Phosphopeptide Enrichment prior to LC-MS/MS

The phosphopeptide enrichment was performed using Fe-IMAC beads (Cell Signaling Technology). The Fe-IMAC beads were washed 3 times with 1mL of wash buffer (0.1% TFA, 80% acetonitrile) prior to the addition of the digested peptides. The beads and digested peptides were incubated on an end-over-end rotator for 30min at room temperature. The peptide-bound beads were then washed 3 times with wash buffer prior to elution with 50uL of elution buffer (50% acetonitrile, 2.5% ammonia). The elution was repeated with 50uL of elution buffer. The eluate was then acidified by the addition of 40 µL of 20% TFA. Following acidification the eluted phosphopeptides were lyophilized to near dryness and were resuspend in 40 µL of 0.1% formic acid.

Nano LC-MS/MS

For Nano LC, the Nanoflow UPLC:Ultimate 3000 nano UHPLC system (Thermo-Fisher Scientific) was used with a trapping column (PepMapC18, 100Å , 100 µm×2 cm, 5µm) and an analytical column (PepMap C18, 100Å, 75 µm×50 cm, 2µm). 10µL sample volume was loaded. The mobile phases were 0.1% formic acid in water (buffer A) and 0.1% formic acid in 80% acetonitrile (buffer B). The flow rate was 250nL/min and a linear gradient was performed from 2 to 8% buffer B over 3 min, from 8% to 20% buffer B over 50 min, from 20% to 40% buffer B over 26 min and from 40% to 90% buffer B in 4 min.

A Q Exactive HF mass spectrometer (Thermo Fisher Scientific, USA) was used with a spray voltage of 2.0kV, capillary temperature of 270°C. The MS parameters were as follows, MS resolution of 70000 at 200 m/z, MS precursor m/z range of 400.0-1800.0, product ion scan range started from m/z 100, CID activation types, 1500.0 min. signal required, 3.00 isolation width, 40.0 normalized collision energy, default charge state of 6, 0.250 activation Q, 30min activation time.

Data dependent MS/MS acquisition was of up to top 20 most intense peptide ions from the preview scan in the Orbitrap.

Data Analysis with Maxquant

The raw MS files were analyzed and searched against human protein database using Maxquant (1.6.2.6). The parameters were set as follows: the protein modifications were carbamidomethylation I (fixed), oxidation (M), Phosphorylation (12phosphor) (variable); the SILAC labels were set to Lys 8 and Arg 10 (heavy); the enzyme specificity was set to trypsin; the maximum missed cleavages were set to 2; the precursor ion mass tolerance was set to 10 ppm, and MS/MS tolerance was 0.6 Da.

Empirical Bayes Moderation Analysis

Individual variances and degrees of freedom for each protein in the experiment was estimated and then this information was combined across all proteins in order to empirically estimate the prior using the experimental data. This prior was then used in a new moderated t-statistic, which allows for correction of instability in the individual variances. The moderated t-statistic was then used as a replacement for the classic linear t-statistic in null hypothesis significance testing. This led to a new set of p-values for each protein and comparison of interest. This was done separately for the phosphoprotein and total protein analysis. After the individual modeling, the phosphopeptides and total peptide results were combined such that the phosphopeptide levels were normalized to the total peptide levels.

Plasmids and Antibodies

HisGb1-STAT1(human 710-750) + CBP TAZ2 (1764-1855) was purchased from Addgene. Site directed mutagenesis was performed to generate the STAT1 S727A version of this plasmid. pSF-NELFB (human full-length) plasmid was a kind gift from Dr. Rong Li (George Washington University). The region encoding amino acids 548-580 was amplified by PCR and subcloned into pGEX-6P1 (GenScript) by using the BamH1 and NotI sites. The region encoding amino acids 520-580 was amplified by PCR and subcloned into pETDUET (Novagen) by using the BamH1 and NotI

sites. Site directed mutagenesis was performed to generate the S557A versions of these plasmids. pCMH6K-TP53BP1 (human full-length) was a kind gift from Dr. Kuniyoshi Iwabuchi (Kanazawa Medical University (1). The region encoding amino acids 215-315 were amplified by PCR and subcloned into pGEX-6P1 by using the BamH1 and NotI sites. Site directed mutagenesis was performed to generate the S265A version of this plasmid. pTRIEX-CUX1-aa612-1328WT and S1237/1270A were kind gifts from Dr. Alain Nepvue (McGill University) (2). pFastBac1 transfer vectors carrying Flag-CDK8WT or KD, HIS-CYCLIN C, HA-MED12-Full length (FL) and CBP-MED13 have been described (3). pDUETKRUP2-HIS-MED12-100WT, was generated by PCR amplification of the region encoding amino acids 1-100 of MED12 contained within pGEX-6P1-MED12aa1-330 described in (4). The amplicon was then subcloned into pRSFDUET-1 vector (Novagen) by using the HindIII and NotI cut sites. pDUETKRUP2-HIS-MED12-100G44D was generated by site directed mutagenesis of the WT plasmid. pGEX2T-Foxk1(mouse)WT and S402/406A which expresses a Foxk1 fragment containing amino acids 380-480 were kinds gifts from Dr. John Blenis (Weill Cornell Medical College) (5). pGEX-6P3-SRSF2aa1-60 was a kind gift from Dr. Beatrice Eymin (Institute For Advanced Biosciences) (6). pGEX6P1-3XCTD was described previously (3).

Rabbit anti CDK8, MED12 and MED13 were from Bethyl laboratories. Rabbit anti CYCLIN C was from BD-Pharmingen. Rabbit anti CDK19 was from Sigma-Aldrich. Rabbit anti FOXK1 and NELFB were from Abcam. Rabbit anti MED4 and MED30 were generated as previously described (7). Rabbit anti β-ACTIN was from Cell Signaling Technology and mouse anti ACTA2 was from Dako. All antibodies were used at a 1:1000 dilution of the stock.

Protein Expression

HIS-Gb1-STAT1aa710-750WT and S727A, GST-NELFBaa548-580WT and S557A, HIS-NELFBaa520-580WT and S557A, GST-TP53BP1aa215-315WT and S265A, HIS-CUX1aa612-1328WT and S1237/1270A, GST-Foxk1aa380-480WT and S402/406A, HIS-MED12aa1-100WT and G44D, GST-SRSF2aa1-60 and GST-CTD were expressed in *E. coli* BL21 (DE3) RIL codon plus cells. The cultures were grown at 37°C until OD_{600} = 0.6 was reached. The cultures were then induced with 0.8mM IPTG and incubated at 20°C for 16hrs.

Flag-CDK8WT and KD, HIS-CYCLIN C, CBP-MED13 and HA-MED12-Full-length (FL)WT and G44D were expressed in High Five insect cells using the baculovirus expression system. Briefly, pFASTBAC1 donor plasmids encoding Flag-CDK8WT, Flag-CDK8KD, HIS-CYCLIN C, CBP-MED13 and HA-MED12-Full-length (FL) were transformed into DH10Bac competent cells (Invitrogen). Following transformation and transposition into bacmid DNAs, colonies containing recombinant bacmids were identified by disruption of the lacZ gene inside the bacmid DNA. The isolated recombinant bacmid DNAs from white, PCR confirmed colonies, were used for transfection of Sf9 insect cells. After three rounds of viral amplification, high-titer baculoviruses were used for infection of High Five insect cells (Invitrogen).

Protein Purification

All HIS-tagged proteins produced in *E. coli* were lysed in 25mM NaH_2PO_4 pH8/200mM NaCl/5mM Imidazole, sonicated and subjected to centrifugation in order to pellet the insoluble fraction. After centrifugation, the soluble fraction was loaded on to a Ni-NTA column (Sigma-Aldrich), washed and eluted by an imidazole gradient from 5 to 200mM imidazole over 5 column volumes. All GST-tagged proteins produced in *E. coli* were lysed in 25mM TRIS pH7.5/200mM NaCl, sonicated and subjected to centrifugation in order to pellet the insoluble fraction. After centrifugation, the soluble fraction was loaded on to a glutathione agarose column (GoldBio), washed and eluted with 10mM reduced glutathione dissolved in 25mM TRIS pH7.5/200mM NaCl The CDK8WT or KD:CYCLIN C complex was purified by co-expressing the proteins in High Five insect cells. The insect cells were lysed in 20 mM Hepes pH 7.9/150 mM NaCl/0.01% NP-40/0.1 mM EDTA/10% glycerol by dounce homogenization. The insoluble fraction was pelleted and the soluble fraction was loaded onto a flag column (Sigma). The column was washed and the CDK8:CYCLIN C complex was eluted with 0.3ug/ul flag peptide (ApexBio). The eluate was dialyzed against 25mM NaH_2PO_4 pH8/200mM NaCl/5mM Imidazole/10mM BME and loaded onto a Ni-NTA column. The Ni-NTA column was washed and the CDK8:CYCLIN C complex was eluted with 25mM NaH_2PO_4 pH8/200mM NaCl/200mM Imidazole/10mM BME.

The MED13:MED12WT or G44D:CDK8WT or KD:CYCLIN C complex was purified by co-expressing the proteins in High Five insect cells. The insect cells were lysed in 20 mM Hepes pH

7.9/150 mM NaCl/0.01% NP-40/0.1 mM EDTA/10% glycerol by dounce homogenization. The insoluble fraction was pelleted and the soluble fraction was loaded onto a pre-washed and equilibrated flag column (Sigma). The column was washed and the complex was eluted with 0.3ug/ul flag peptide (ApexBio). The eluate was then further purified by a glycerol gradient from 10 to 30% glycerol.

In-vitro Kinase Assay

Purified Flag-CDK8WT or KD:HIS-CYCLIN C complexes were complexed with HIS-MED12aa1-100WT or G44D for 1 hr at 20°C. 40ng of either Flag-CDK8:HIS-CYCLIN C:HIS-MED1:100 or 400ng of CBP-MED13:HA-MED12-FL:Flag-CDK8:HIS-CYCLIN C complex was incubated with purified substrate (1-4ug) in kinase reaction buffer (25 mM Tris, pH 7.5, 20mM MgCl2), 2.5 µCi of [γ-^{32}P]ATP. Reactions were incubated for 30 min at 30°C, stopped with the addition of Laemmli sample buffer, processed by SDS-PAGE, stained with Coomassie and visualized by PhosphorImager analysis.

Quantitative Real-time PCR

Myometrium side population cells (MM SP) cells in exponential growth were treated with DMSO or 100nM CCT251545 for 3hr. RNA was extracted from treated cells with the Quick RNA-miniprep kit (Zymo). RNA was then reverse transcribed with the Reverse Transcription System from Promega using OligodT. Quantitative real-time PCR (qRT-PCR) was performed using SSO Advance SYBR Green PCR Mix from Biorad. Primer pair specificity was determined by generation of a single peak for dissociation curve (melting curve) at the end of RT-PCR cycling program. GAPDH was used for endogenous gene controls. All primers used in this study were designed by Primer Blast of NCBI and synthesized by Sigma. Primer sequences are listed in Table S5.

RNA-sequencing

MM SP cell treatment and RNA preparation was as described for quantitative real-time PCR experiments. RNA library prep and sequencing (on an Illumina HiSeq 2000 or 3000with >32 million reads/sample) was performed at the UT Health San Antonio Genomics Sequencing

Facility. Sequencing was done on triplicate independent experiments. All samples were aligned to the UCSC human genome build hg19 using TopHat2, and bam files from the alignment were processed using HTSeq-count to obtain the read counts per gene in all samples. Analysis of the read count data was performed using R version 3.6.3. Raw count data was first filtered to include genes with minimum CPM 1 in at least 1 library across all samples. The data was then transformed for visual inspection using regularized log2 transformation function of R package DESeq2 version 1.32. Visual inspection was done by performing principal component analysis and by generating a Pearson's correlation heat map using basic R functions, R package heatmap and custom R scripts. Data normalization and differential expression (DE) were performed using R package DESeq2, version 1.32. P values of the DE analysis were adjusted for multiple testing using Benjamini-Hochberg procedure. DE analysis results were annotated with gene biotypes, gene descriptions, and Ensembl identifiers using R package biomaRt. Statistically and biologically significant differentially expressed genes (DEGs) were defined with the following criteria: adjusted p-values <0.05 and absolute Fold change ≥1.5.

Gene set enrichment and gene ontology analysis

Gene set enrichment analysis (GSEA) was performed using expressed genes, with CPM (averaged over all cells) > 1, ranked based on average fold-change in vehicle compared to treatment with the mediator kinase inhibitor (CCT251545). Significantly enriched Hallmark (MSigDB v7.1) gene sets were identified using GSEA pre-ranked software (v. 4.0.3) with default settings. The Hallmark gene sets with $p < 0.05$ were identified as significantly enriched. Gene ontology (GO) analyses were performed using the DAVID functional annotation tool.

Transcription Factor Binding Site Enrichment Analysis

Transcription factor binding site enrichment analysis was performed using CiiiDER software (8). Transcription factor binding sites were searched against the JASPARCORE2020 matrices with 0.1 threshold deficit. Binding sites were searched ±2kb from the transcriptional start site. Binding sites that were enriched with a gene coverage p-value < 0.05 were considered significant.

Supplemental References

1. Iwabuchi K, Li B, Massa HF, Trask BJ, Date T, Fields S. Stimulation of p53-mediated transcriptional activation by the p53-binding proteins, 53BP1 and 53BP2. J Biol Chem 1998;273:26061-8.
2. Sansregret L, Gallo D, Santaguida M, Leduy L, Harada R, Nepveu A. Hyperphosphorylation by cyclin B/CDK1 in mitosis resets CUX1 DNA binding clock at each cell cycle. J Biol Chem 2010;285:32834-43.
3. Zhou H, Spaeth JM, Kim NH, Xu X, Friez MJ, Schwartz CE *et al.* MED12 mutations link intellectual disability syndromes with dysregulated GLI3-dependent Sonic Hedgehog signaling. Proceedings of the National Academy of Sciences of the United States of America 2012;109:19763-8.
4. Park MJ, Shen H, Spaeth JM, Tolvanen JH, Failor C, Knudtson JF *et al.* Oncogenic exon 2 mutations in Mediator subunit MED12 disrupt allosteric activation of Cyclin C-CDK8/19. J Biol Chem 2018.
5. He L, Gomes AP, Wang X, Yoon SO, Lee G, Nagiec MJ *et al.* mTORC1 Promotes Metabolic Reprogramming by the Suppression of GSK3-Dependent Foxk1 Phosphorylation. Mol Cell 2018;70:949-60 e4.
6. Edmond V, Moysan E, Khochbin S, Matthias P, Brambilla C, Brambilla E *et al.* Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. EMBO J 2011;30:510-23.
7. Kim S, Xu X, Hecht A, Boyer TG. Mediator is a transducer of Wnt/beta-catenin signaling. J Biol Chem 2006;281:14066-75. Epub 2006 Mar 24.
8. Gearing LJ, Cumming HE, Chapman R, Finkel AM, Woodhouse IB, Luu K *et al.* CiiiDER: A tool for predicting and analysing transcription factor binding sites. PLoS One 2019;14:e0215495.