NORTHWESTERN UNIVERSITY

Enhancer Malfunction Underlies Transcriptional Dysregulation in Uterine Leiomyoma

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Health and Life Sciences

Driskill Graduate Program in Life Sciences (DGP)

By

Mthabisi Bongani Moyo

EVANSTON, ILLINOIS

June 2020

ABSTRACT

Enhancer Malfunction Underlies Transcriptional Dysregulation in Uterine Leiomyoma

Mthabisi Bongani Moyo

Uterine leiomyomas (fibroids) are a major source of gynaecologic morbidity in reproductive age women and are characterised by the excessive deposition of a disorganised extracellular matrix, resulting in rigid benign tumours. Clinically, leiomyoma patients usually present with pelvic pain, urinary incontinence, as well as heavy cyclic and non-cyclic bleeding. Curative treatment options are limited, with hysterectomies currently the only guaranteed treatment to prevent reoccurrence.

Uterine fibroid tumours are composed primarily of smooth muscle cells and fibroblasts. They have also been shown to be clonal, with multiple tumours from the same patient sometimes originating from different tumour initiating cells. Leiomyomas fall into four main mutational subtypes: tumours with genetic aberrations in either fumarate hydratase (*FH*), mediator of transcription subunit 12 (*MED12*), high mobility group AT-hook 2 (*HMGA2*), or type IV collagens. Somatic mutations in Mediator complex subunit *MED12* have been implicated as the causal genetic lesion in the majority of leiomyoma. The functional consequence of this mutation with respect to altered epigenetic events and subsequent transcriptional dysregulation remains poorly understood.

Epigenetic control of gene expression by promoter-distal *cis*-regulatory elements such as enhancers is an established and important mechanism of gene regulation, with aberrant epigenetic changes at distal sites being implicated in human disease. Despite this, the role of enhancer regulation in uterine leiomyomas remains unexplored. In this study, optimised

extraction procedures for fibrotic tissues, coupled with high resolution ChIP-sequencing, promoter capture Hi-C, and RNA-sequencing of matched normal myometrium and diseased leiomyoma tissues, are used to provide an integrative analysis of transcriptional dysregulation mechanisms in leiomyomas. Through the characterisation of the first chromatin interactome in uterine leiomyomas, this study demonstrates that modified enhancer architecture is an important mechanism of transcriptional dysregulation in uterine fibroids.

Activator protein 1 (AP-1) is a dimeric transcription factor that responds to varied stimuli and is responsible for the regulation of many cellular processes such as proliferation and differentiation. AP-1 subunits belonging to *JUN*, *FOS* and *ATF* gene families have previously been demonstrated to be down regulated in uterine leiomyomas. This study highlights the link between decreased AP-1 subunit gene expression and modifications in leiomyoma tissue enhancer architecture. The loss of AP-1 expression results in the depletion of AP-1 occupancy on chromatin in leiomyoma tissue samples, which correlates with changes in H3K27 acetylation at distal *cis*-regulatory elements.

Importantly, this study also demonstrates that silencing of AP-1 subunits in primary human uterine smooth muscle cells leads to large-scale transcriptional dysregulation. In addition, loss of AP-1 causes significant alteration to the H3K27Ac cistrome in uterine smooth muscle cells, primarily at promoter-distal sites, thereby partially recapitulating epigenetic changes seen in leiomyoma tissue samples. The data supports the hypothesis that AP-1 plays an important mechanistic role in enhancer maintenance and the loss of AP-1 results in altered enhancer architecture in uterine smooth muscle cells. These findings establish AP-1 driven aberrant enhancer regulation as an important mechanism of leiomyoma disease pathogenesis.

ACKNOWLEDGMENTS

"Umuntu ngumuntu ngabantu." ("A person is a person through other people."; Nguni proverb)

Thank you to everyone who contributed, directly and indirectly, to the completion of this work.

LIST OF ABBREVIATIONS

ADAM	A disintegrin and metalloproteinase
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
BMP	Bone morphogenetic protein
BRE	TFIIB recognition element
Cas9	CRISPR associated protein 9
CDK	Cyclin dependent kinase
CGI	CpG island
ChIP	Chromatin immunoprecipitation
CHiC	Capture Hi-C
CCR	Complex chromosomal rearrangement
CpG	Cytosine-guanine dinucleotide
CRE	*Cis*-regulatory DNA element
CRISPR	Clustered regularly interspaced short palindromic repeats
DPE	Downstream core promoter element
DSIF	(5,6-dichloro-1-β-d-ribofuranosylbenzimidazole) sensitivity-inducing factor
ECM	Extracellular matrix
EGF	Epidermal growth factor
ERK	Extracellular signal regulated kinase
FGF	Fibroblast growth factor
FH	Fumarate hydratase

FIGO	International Federation of Gynaecology and Obstetrics
FSH	Follicle stimulating hormone
GAG	Glycosaminoglycans
GDF	Growth and differentiation factor
GTF	General transcription factor
G6PD	Glucose-6-phosphate dehydrogenase
GnRH	Gonadotropin-releasing hormone
HDR	Homology directed recombination
HIF	Hypoxia inducible factor
HLRCC	Hereditary leiomyomatosis and renal cell cancer
HMG	High mobility group
HMGA2	High mobility group AT-hook 2
HR	Homologous recombination
HUMARA	Human androgen receptor gene assay
HUtSMC	Human uterine smooth muscle cells
IGF	Insulin-like growth factor
Inr	Initiator
IRS4	Insulin receptor substrate 4
JNK	JUN N-terminal kinase
KDM2B	Lysine-specific demethylase 2B
LH	Luteinising hormone
LNG-IUS	Levonorgestrel-releasing intrauterine devices

MAPK	Mitogen activated protein kinase
MED12	Mediator of transcription subunit 12
MMP	Matrix metalloproteinases
NELF	Negative elongation factor
NFR	Nucleosome free region
NHEJ	Non-homologous end joining
NOD-SCID	Non-obese diabetic-severe combined immune deficiency
NSAID	Non-steroidal anti-inflammatory drug
PALM-COEIN	Polyps; Adenomyosis; Leiomyomas; Malignancy and hyperplasia; Coagulopathy; Ovulatory dysfunction; Endometrial; Iatrogenic; Not classified
PARP	Poly(ADP)-ribose polymerase
PDGF	Platelet-derived growth factor
PG	Proteoglycan
PI3K	Phosphoinositide 3-kinase
PIC	Pre-initiation complex
PPROM	Preterm premature rupture of the membranes
P-TEFb	Positive transcription elongation factor-b
PTM	Posttranslational modification
R-Smad	Regulatory Smad
RNAP	Ribonucleic acid polymerase
SERMs	Selective oestrogen receptor modulators

SLRP	Small leucine-rich proteoglycan
SPRMs	Selective progesterone receptor modulators
TAF	TBP-associated factors
TBP	TATA binding protein
TGF-β	Transforming growth factor-β
VEGF	Vascular endothelial growth factor

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1. INTRODUCTION

1.1. Uterine Leiomyoma

Uterine leiomyomas (fibroids) are very common benign neoplasms that originate from the smooth muscle layer of the uterine wall, the myometrium. Fibroids affect reproductive age women, with no reported cases in children, very few reported cases in adolescents, and a decreased incidence post menopause [1-3]. In addition to the uterus, leiomyomas also occur in the smooth muscle layers of the bladder, oesophagus, stomach, blood vessels and the skin, albeit at significantly lower incidence rates [4-9]. Preoperatively, uterine leiomyomas are morphologically very similar to uterine leiomyosarcomas, which are rare malignant tumours also originating from the myometrium [10-12]. In addition, signs and symptoms associated with uterine leiomyomas and leiomyosarcomas are usually indistinguishable, with the exception that leiomyosarcomas show consistent tumour growth in postmenopausal women. However, leiomyomas and leiomyosarcomas are histopathologically distinct, although some cases involving atypical leiomyomas also prove difficult to distinguish from leiomyosarcomas. A common disease aetiology may exist as leiomyosarcomas are often found in patients that also have leiomyomas. In some cases, leiomyomas do develop into sarcomas, although very rarely [12].

Uterine leiomyomas are classified according to the International Federation of Gynaecology and Obstetrics (FIGO) classification system for causes of abnormal uterine bleeding in reproductive-age women [13]. The system classifies causes of abnormal uterine bleeding into 9 categories (PALM-COEIN) based on structural features that can be measured visually: polyps; adenomyosis; leiomyomas; malignancy and hyperplasia (i.e. PALM) and non-structural features that cannot be measured visually: coagulopathy; ovulatory dysfunction; endometrial; iatrogenic; not classified (i.e. COEIN). Under this primary system of classification,

a secondary classification of leiomyomas exists that considers the location of the tumour based on whether they grow within or outside of the uterine wall (intramural or pedunculated) and whether they project into or away from the uterine cavity (submucosal or subserosal).

Data on leiomyoma incidence by race reveals that approximately 70% of all reproductive age White women develop fibroid tumours by age 50, with the cumulative incidence rising to over 80% in reproductive age Black women [14]. Results from Marshall et al. suggest similar incidence rates between Asian women and White women, with slightly higher rates in Latino women as compared to White women [15]. In addition to a higher prevalence, Black women also have earlier onset uterine leiomyomas and higher morbidity, which includes more tumours [14,16,17]. The causes of the observed racial disparity in leiomyoma disease are currently unknown.

Despite a high prevalence of uterine fibroids, it is notable that a significant number of cases are possibly asymptomatic [16,18]. A study of 101 Black and White women between the ages of 18 – 30 years who did not report clinical symptoms associated with fibroids found that 26% and 7% Black and White women respectively had previously undiagnosed fibroid tumours [19]. Given the relatively small size of the tumours identified and the relatively young ages of the women in the study, it is likely that the tumours constitute early-stage leiomyomas that are likely to cause symptoms in patients over time. Other studies involving women between the ages of 25 – 56 years of age have investigated incidence rates of asymptomatic women with leiomyomas [20-22]. These studies noted overall prevalence rates of 9 – 16%, with significant variation depending on race and age.

Profuse menstrual bleeding (menorrhagia) is the most common symptom associated with leiomyoma disease [1,16,23]. Prolonged bleeding during menstrual periods and bleeding between

periods have also been reported in women with uterine fibroids, with some women developing anaemia as a consequence of menorrhagia. Pain during sexual intercourse (dyspareunia) and non-cyclic pelvic pain are also symptoms associated with fibroids. Although dysmenorrhea (menstrual cramps) has also been noted in women diagnosed with uterine leiomyomas, there is no observable difference in incidence rates between women diagnosed with uterine fibroids and those that did not have fibroids [24]. Urinary incontinence characterised by increased frequency, urgency and nocturia have all been reported in women with leiomyomas [1,25].

Reproductive outcomes are also negatively impacted by the development of leiomyomas, although this depends on the location of the tumours. Systematic reviews of studies related to reproductive outcomes in women with uterine leiomyomas concluded that a decrease in fertility and a higher risk of miscarriage are associated with leiomyomas, with submucosal tumours having a greater effect on fertility [26-28]. Fibroids were also found to have adverse effects on pregnancy and are associated with an increased risk of preterm delivery, attachment of the placenta near the cervical opening (placenta praevia) and foetal malpresentation, leading to a higher rate of caesarean delivery. Other pregnancy related symptoms associated with fibroids include first trimester bleeding, placental abruption, preterm premature rupture of the membranes (PPROM), postpartum haemorrhage, and postpartum inflammation of the endometrium and myometrium (endomyometritis).

Several mechanisms have been proposed for leiomyoma-related adverse reproductive outcomes, with anatomical alterations of the uterus and uterine cavity as a result of leiomyomas being suggested as a major cause of decreased embryo implantation [28,29]. Abnormal uterine contractility, alterations in growth factor and hormonal balance in the leiomyomatous uterus

have also been implicated in abnormal endometrial receptivity, which leads to decreased fertility rates [30].

1.2. Leiomyoma treatment options

Multiple surgical and non surgical options are available for the treatment of uterine fibroids [1]. As leiomyomas are hormone responsive, a majority of drug treatment options for leiomyomas target their hormone response pathways. Selective progesterone receptor modulators (SPRMs) such as Ulipristal Acetate have been shown to be effective in fibroid tumour shrinkage and managing fibroid-related abnormal uterine bleeding [31]. SPRMs act by modulating the ability of the progesterone receptor to recruit and interact with its coregulators as well as affecting posttranslational modifications of the receptor and its coregulators [32]. In uterine leiomyomas, this results in the decreased expression of angiogenic growth factors, thereby decreasing vascularisation of the tumour [33]. In addition, SPRMs have anti-proliferative effects by increasing caspase-3 (*CASP3*) gene expression and decreasing *BCL-2* expression. Studies have demonstrated that repeated 12-week treatments with Ulipristal Acetate are an effective long-term management strategy of leiomyoma symptoms [34]. In addition, the drug has been shown to be effective at tumour reduction as a pre-surgery treatment option [35]. Despite the success of Ulipristal Acetate, other SPRMs suitable for the treatment of fibroid symptoms have not passed clinical trials. Although effective at tumour reduction and control of leiomyoma-associated abnormal uterine bleeding, morphological changes to endometrial tissue as a direct result of SPRMs such as Asoprisnil has hindered progress towards viable alternatives to Ulipristal Acetate [36,37].

Contraceptives targeting the progesterone hormone signalling pathway, in particular

Levonorgestrel-releasing intrauterine devices (LNG-IUS), are used as a treatment for fibroid-related abnormal bleeding [38-40]. Although the mechanism of action resulting in decreased menorrhagia is unclear, Levonorgestrel, a progestin, acts as a progesterone receptor agonist and causes rapid decidualization of the endometrial stroma and inhibition of endometrium proliferation, resulting in a thinning of the endometrium and a significant reduction in blood loss [41-43]. In addition, Levonorgestrel causes changes to uterine vascular morphology and affects blood vessel integrity in the endometrium. As a result, Levonorgestrel-releasing intrauterine devices have become a mainstay first-line therapy for leiomyoma-associated abnormal uterine bleeding.

Selective oestrogen receptor modulators (SERMs) have also been used as a possible treatment for uterine fibroids. Preclinical studies in cell lines have demonstrated an oestrogen-dependent growth of leiomyoma cells and the inhibitory effect of SERMs such as Raloxifene and Tamoxifen on leiomyoma cell growth [44,45]. A 22.2% difference in tumour size was observed between 13 premenopausal women with fibroids treated with Raloxifene for 3 months as compared to a control group of 12 premenopausal women, also with fibroids but receiving placebo treatment [46]. In studies of postmenopausal women, Raloxifene was found to have a significant effect on tumour size, with a 17.9% decrease in leiomyoma tumour volume after 3 months of treatment as compared to no change in the placebo group [47]. Although approved for the treatment of osteoporosis in postmenopausal women, larger studies to determine efficacy of Raloxifene in uterine fibroid treatment are yet to be performed.

Aromatase inhibitors, a class of anti-oestrogenic therapeutics, have also demonstrated efficacy at fibroid tumour shrinkage. Aromatase cytochrome P450 ($P450_{arom}$) catalyses the

conversion of androgenic steroids to oestrogen by the hydroxylation and elimination, as formic acid, of the androgenic substrate's methyl group at the C19 position and subsequent aromatisation of the substrate's A-ring [48,49]. Aromatase inhibitors competitively and reversibly bind to the steroid-binding site of the enzyme and act as a ligand for the haem group of the enzyme. A significant reduction in tumour volume was observed in small studies of premenopausal women with leiomyomas treated with the aromatase inhibitors Anastrazole or Letrozole [50-53]. Despite encouraging initial studies, larger clinical trials are yet to be performed to establish the efficacy of aromatase inhibitors in the treatment of uterine fibroids.

Gonadotropins are glycoprotein hormones produced in the gonadotrope cells of the anterior pituitary. Examples of gonadotropins include follicle stimulating hormone (FSH) and luteinising hormone (LH). Through the regulation of sex steroid hormone production in the gonads, LH and FSH are responsible for growth, reproductive development and fertility in mammals [54]. Their secretion is controlled by the gonadotropin-releasing hormone (GnRH), which is itself secreted by GnRH neurons in the hypothalamus and activates gonadotropin-releasing hormone receptors (GnRHR) in gonadotrope cells of the pituitary, thereby stimulating gonadotropin synthesis [55,56]. GnRH agonists and antagonists have proven to be effective therapies in the reduction of fibroid tumour volume. GnRH agonists such as Leuprolide Acetate, Goserelin, and Nafarelin inhibit gonadotropin secretion by binding to GnRH receptors with a higher binding affinity than GnRH and inducing a short-lived production and hyper secretion of LH and FSH [56]. Unlike GnRH, which has a short half-life of under 5min, chronic activation by GnRH agonists results in the desensitisation of GnRH receptors and a concomitant decrease in secretion of gonadotropins. A reversible gonadotropin-driven decrease in oestrogen and

progesterone levels negatively impacts the growth of sex steroid-dependent leiomyomas leading to a decrease in tumour size [57].

GnRH antagonists competitively occupy the GnRH receptor binding site and thereby inhibit production and secretion of LH and FSH [58]. Many studies have demonstrated the efficacy of GnRH antagonists in fibroid tumour reduction and show that GnRH antagonists achieve a more rapid rate of tumour volume reduction than GnRH agonists [59,60]. Despite the commercial availability of GnRH antagonists, none have received approval as a treatment for uterine fibroids, although clinical trials for potential efficacious GnRH antagonists are ongoing [1,61]. Although the primary mechanism of GnRH agonist/ antagonist action in leiomyomas is by inhibiting sex steroid hormone production and secretion, studies have demonstrated a direct inhibition of extracellular matrix associated gene expression in leiomyoma cells [62,63].

In addition to SPRMs, SERMs, aromatase inhibitors and GnRH agonists and antagonist, other drugs such as anti-fibrinolytics and non-steroidal anti-inflammatory drugs (NSAIDs) have also been used to treat fibrinolysis and pain associated with fibroids respectively [1]. Currently, all available non surgical treatment options for uterine leiomyomas are non curative and a significant number of patients ultimately undergo surgical procedures to treat fibroids [64,65]. Surgical options for treatment of uterine leiomyomas range from minimalist uterine-sparing therapies to more drastic therapies that require removal of the uterus (hysterectomy).

The most preferred uterine-sparing therapy is a myomectomy. This involves the removal of the fibroid tumour/s alone and is notable for its fertility-preserving benefits as compared to a hysterectomy. In addition, studies demonstrate higher fertility rates in women who had been diagnosed with leiomyomas and underwent a myomectomy than those who did not [66,67]. A

myomectomy can be performed by the transvaginal insertion of a resectoscope into the uterine cavity, an approach that is appropriate for pedunculated fibroids localised in the uterine cavity or submucosal fibroids with < 50% of the fibroid in an intramural location [1,68]. For submucosal fibroids with > 50% of the fibroid in an intramural location and subserosal fibroids, laparoscopic myomectomy and abdominal myomectomy are the common uterine-sparing approaches used, with laparoscopic procedures being the preferred method as it results in less blood loss and post operative pain [69-71].

In addition to myomectomies, uterine artery embolisation and MRI-guided focused ultrasound surgery are two recent uterine-sparing treatment options that are available to uterine fibroid patients. MRI-guided focused ultrasound surgery uses ultrasonic energy to cause molecular vibrations in tissue, thereby heating fibroid tumours to temperatures between 60 – 90°C, resulting in cell death and tissue necrosis [72]. This approach has been shown to be effective at sustained tumour shrinkage and amelioration of leiomyoma-related symptoms whilst maintaining reasonable conception rates and positive pregnancy outcomes as early as 8 months post surgery [73-75]. Uterine artery embolisation aims to limit blood supply and nutrients to the tumour. However, unlike other uterine-sparing therapies, this procedure does not preserve fertility and is more appropriate for women who do not wish to bear children but would prefer to not have a hysterectomy [76-78].

Despite their efficacy in treating leiomyomas, uterine-sparing therapies do not prevent reoccurrence of tumours and as such are not considered curative treatments. Therefore, hysterectomies are often performed and are the only curative treatment available for uterine fibroids. Similar to myomectomies, hysterectomies can be performed vaginally, laparoscopically

or through the abdomen [79]. Although hysterectomies are a very effective curative treatment, the resulting financial burden is high and result in irreversible loss of fertility [80]. Given the large number of fibroid-related hysterectomies performed and the projected rise in leiomyoma-related inpatient care in the coming decades, there is a need for more effective non-surgical treatment options [64,65].

1.3. Cellular and molecular composition of uterine leiomyomas

Studies have demonstrated that although uterine leiomyomas are composed of distinct smooth muscle and fibroblast cell populations, these are derived from a single cell that differentiates into the various subpopulations. Early studies took advantage of the X-linked gene Glucose-6-phosphate dehydrogenase (G6PD), which has two isoforms, to determine the clonality of fibroid tumours [81]. Random X-inactivation ensures that only one isoform will be expressed in each cell of women that are heterozygous for G6PD isoforms. Accordingly, tissue samples taken from the myometrium showed an equal representation of both isoforms, whereas the majority of samples taken from individual fibroid tumours show expression of only one isoform in any given tumour, demonstrating the clonal origin of each tumour. More recent techniques have relied on differences in methylation status of active and inactive X chromosome alleles. The human androgen receptor gene assay (HUMARA) utilises the methylation-sensitive restriction enzyme *HhaI* to assay the differential methylation status of the exon 1 CAG repeat polymorphism in the X-linked human androgen receptor gene, with only the methylated inactive allele being amplified [82]. Myometrium samples showed random amplification of CAG repeats from both alleles whereas leiomyoma samples showed non-random amplification of CAG repeats from one allele, further demonstrating the monoclonality of fibroid tumours [83]. Interestingly, in women

with multiple leiomyomas, whilst each tumour was found to be monoclonal, the individual tumours in a significant number of patients were found to express different isoforms of *AR* [84]. This demonstrates not only the monoclonality of uterine leiomyomas but also that each clonal tumour originates from different tumour-initiating cells.

The HUMARA assay has also been used to delineate the pathogenic relationship between leiomyomas and leiomyosarcomas [85]. Zhang et al. determined that leiomyosarcomas are also clonal in origin and individual clonal tumours, similarly to leiomyomas, originate from different initiating cells. Also, in patients that developed leiomyomas as well as leiomyosarcomas, adjacent tumours had similar X-inactivation profiles. Furthermore, a morphological transition from a leiomyoma phenotype to a leiomyosarcoma phenotype in a patient with a leiomyoma adjacent to a leiomyosarcoma was observed. This evidence suggests that although distinct diseases, leiomyomas may share common developmental origins and benign leiomyomas do possess the potential to evolve into malignant leiomyosarcomas.

Uterine leiomyomas are composed of four main cell populations: $ALDH^-/CD90^-$ vascular smooth muscle cells; $ALDH^-/CD90^+$ smooth muscle cells; $ALDH^+/CD90^+$ fibroblasts; and $ALDH^-/CD90^{+bright}$ fibroid-associated fibroblasts [86]. Furthermore, Holdsworth-Carson et al. also determined that the individual fibroblast and smooth muscle cell populations all have identical X-inactivation profiles thereby suggesting that all the tumour subpopulations originated from the same tumour-initiating cell, which then differentiated into the various subpopulations seen in leiomyomas. In addition to the four main subpopulations, a side population of cells has also been identified [87,88]. The side population, which constitutes < 1% of cells found in a given tumour, where shown to be multipotent, suggesting that they are possibly tumour-initiating cells

[87]. Cell lines generated from the side population of cells were tumour forming when single cells were injected into the left-flank of NOD-SCID mice (i.e. non obese diabetic mice with a spontaneous mutation in the DNA-dependent protein kinase Prkdc, resulting in severe combined immune deficiency) [87]. They were similarly found to be tumour forming when xenografted into mice as kidney capsules.

Leiomyoma cells have also been shown to have a low mitotic index, although they do have a relatively higher index than matched myometrium cells [89]. Paradoxically, despite a low mitotic index, leiomyomas have also been demonstrated to have a high growth rate, with documented increases in size of greater than 20% within a six-month period for 38% of all tumours surveyed [17]. This is explained primarily by a disease pathogenesis involving excessive deposition of extracellular matrix components [90].

1.4. Extracellular matrix composition in uterine leiomyomas

The extracellular matrix (ECM) is a network of cross-linked extracellular macromolecules that function as cellular scaffolds and are important for maintaining tissue structure integrity [91]. The ECM is also involved in cell adhesion and de-adhesion, and in conjunction with the cytoskeleton plays a crucial role in cell locomotion and migration. Another key characteristic of the extracellular matrix is its mechano-transduction properties. Its ability to sense and relay mechanical signals to cells is thought to be important in cell fate decision making [92]. In addition, it is also involved in many diverse cellular functions such as regulating cell growth, proliferation, survival, and apoptosis [92-94]. Given the diverse and important roles fulfilled by the extracellular matrix, aberrant expression, secretion or degradation of ECM proteins has been shown to result in the development of multiple disease pathologies such as tissue fibrosis, pro-inflammatory

diseases and various cancers [95-101]. The extracellular matrix is composed of fibrous proteins, glycoproteins and proteoglycans, which are actively secreted by the cell into the extracellular space. The main ECM fibrous proteins are collagens, fibronectin, elastins, and laminins, with collagens being the most abundant and the main structural protein in the extracellular matrix [91].

Collagens are a superfamily comprised of 28 distinct subtypes and the high tensile strength of triple-helical hydrogen-bonded α-chains make them ideally suited as the main structural protein of the ECM [92,102]. Fibroblasts are responsible for the majority of collagen synthesis and secretion. They also play an important role in the organisation of matrix collagen fibrils into sheets and cables [103]. Collagens can be classified into seven functional categories: (i) fibrillar collagens which provide tissues with tensile strength and elasticity; (ii) network-forming collagens that provide cell anchorage and molecular filtration; (iii) relatively short fibril-associated collagens with interrupted triple helices (FACITs), which are localised in basement membranes and link fibrillar collagens with other ECM molecules; (iv) membrane-associated collagens with interrupted triple helices (MACITs) that act as cell adhesion-regulating surface markers; (v) anchoring fibrils which are essential for the integrity of the dermoepidermal junction; (vi) tetrameric beaded filament-forming collagens which form structural links with cells; (vii) and endostatin-producing collagens which are important for regulating endothelial cell migration and angiogenesis through the release of endostatins [102,103]. Elevated collagen expression of multiple subtypes has been experimentally determined in uterine leiomyomas [90,104,105]. This widespread aberrant expression of collagens is a major characteristic of fibroid morphology and pathology.

Fibronectin (FN) is a highly and ubiquitously expressed dimeric ECM protein consisting

of disulphide-bonded subunits, each composed of three types of repeating units, type I, II, III repeats, and is important for tissue formation, remodelling and repair [103]. Alternative splicing of the singular *FN1* gene results in multiple protein variants and can be classified, based on its solubility, as soluble plasma FN (pFN) and insoluble cellular FN (cFN). Through fibronectin matrix assembly, FN is arranged into an interconnected network of long super molecular fibres that mediate cell attachment and extracellular matrix interaction [106]. It also has multiple binding domains for ECM proteins that include collagens, heparin, fibrin, and integrins [103,106]. Dysregulation of fibronectin has prominently been identified in disease states and plays a role in tumour progression and fibrosis [103]. Significantly, FN is upregulated in uterine leiomyomas [107].

Another major group of fibrous proteins are elastins, which are insoluble hydrophobic proteins that are tightly associated with collagen fibrils and are important for tissue elasticity [91]. Tropoelastins, the soluble precursor of elastin molecules, are secreted by the cell and cross-linked to each other. Similar to fibronectin, a single elastin gene (*ELN*) is alternatively spliced, giving rise to multiple tropoelastin isoforms [108]. Elastins associate with fibrillin-rich microfibrils to form elastin fibres, which are deposited on pericellular microfibril bundles [109]. In addition to fibrillins, microfibril associated proteins (MFAP1-5), Elastin microfibril interface located proteins (EMILIN1-3), and fibulins (FBLN1-5) have also been shown to play an integral role in elastin function as critical components of microfibril bundles [103,109,110]. Limited studies on elastin composition have been performed in fibroids. However, Wolańska et al. reported increased elastin levels in leiomyomas as compared to myometrium [111].

Laminins are major basement membrane heterotrimeric glycoproteins that play a role in cell differentiation, adhesion, extracellular matrix organisation, as well as the maintenance of cell

shape and movement [103,112]. They are critical components of matrix receptor-mediated basement

membrane assembly and their secretion into the extracellular space is necessary for the

incorporation of other key molecules into the basement membrane [112]. Fifteen laminin proteins

have been identified (Laminin-111 – Laminin-523), each composed of one α-chain encoded by

LAMA1-5 genes, one β-chain encoded by *LAMB1-4* genes, and one γ-chain encoded by *LAMC1-3* genes. Laminins interact with each other via their laminin N-terminal domains to form super

molecular structures in the basement membrane. In addition, they also interact with collagens,

fibronectin and various other ECM glycoproteins and proteoglycans primarily through their C-terminus LG domains [103]. Studies have suggested that dysregulation of laminins in may be

involved in uterine leiomyoma pathogenesis [113]. Malik et al. report increased expression of

laminins in leiomyoma cells and suggest altered laminin-dependent mechanical signalling in

leiomyomas.

Proteoglycans (PGs) are proteins covalently bonded with glycosaminoglycans (GAGs),

which are linear acidic disaccharide repeats composed of *N*-acetylated hexosamines and D-/L-hexuronic acid [103]. In addition to fibrous proteins such as collagens, proteoglycans play a crucial

role in the extracellular matrix and are involved in cell proliferation, adhesion, migration,

differentiation, and cell signalling [103,114]. Six types of glycosaminoglycans have been identified,

namely chondroitin sulphate (CS), dermatan sulphate (DS), heparan sulphate, heparin, keratan

sulphate (KS), and hyaluronic acid. Extracellular secreted proteoglycans interact with ECM

constituents primarily through their GAG side chains and based on the side chains, most

proteoglycans can be classified into two main groups: hyaluronan-binding PGs (hyalectans) and

small leucine-rich proteoglycans (SLRPs). Proteoglycans have been implicated in many diseases

including cancer proliferation and metastasis, as well as chronic inflammatory disorders such as atherosclerosis, renal disorders and fibrosis [101,115-119]. Relevantly, dysregulation of proteoglycan expression has been implicated in uterine fibroid pathogenesis, with sulphated proteoglycans being notably increased in leiomyomas [105,111,120]. In particular, versican, decorin, and fibromodulin, have all been shown to be dysregulated in leiomyomas [121,122].

1.5. Dysregulation of extracellular matrix-associated proteins in uterine leiomyomas

In addition to core extracellular matrix proteins such as fibrous proteins and proteoglycans, bioinformatics and proteomics-based approaches have been used to identify and classify additional extracellular matrix components. This full complement of core ECM and ECM-associated proteins has been defined as the matrisome [123]. Mass spectrometric analysis of tissue samples sequentially extracted to remove cytosolic proteins, nuclear proteins, membrane proteins, and cytoskeletal proteins, was used to identify 278 core matrisome proteins (i.e. 243 fibrous proteins and 35 proteoglycans) and 778 ECM-associated proteins consisting of ECM remodelling enzymes and secreted factors, which are important for the regulation of the highly dynamic extracellular matrix that is constantly undergoing degradation and remodelling [124].

The most relevant group of ECM-associated proteins is the proteolytic metalloproteinases, which are responsible for ECM degradation. Three main classes of metalloproteinases have been identified, namely matrix metalloproteinases (MMPs), the transmembrane a disintegrin and metalloproteinase (ADAMs), and a disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) [94,103]. Metalloproteases target a diverse range of ECM proteins, with different proteases having substrate preferences for different ECM components. They are usually secreted as inactive precursors to prevent

uncontrolled ECM degradation. The activity of the different families of metalloproteases is regulated by tissue inhibitors of metalloproteinases (TIMPs) [94]. In addition to proteolytic enzymes, lysyl oxidases (LOX) and lysyl hydroxylases are also important for ECM remodelling and regulate the level of cross-linking between ECM components thereby modulating ECM stiffness [123,125].

The extracellular matrix also functions as a ligand reservoir, sequestering growth factors as well as other signalling molecules [123,126]. Growth factors are specialised polypeptide molecules involved in intercellular signalling and act over short distances to promote cell proliferation, motility, apoptosis and differentiation [126,127]. They are bound to ECM proteins and are also considered to be constituents of the matrisome [123]. Importantly, altered expression and secretion of growth factors has been well documented in uterine leiomyomas [1,122,128]. Dysregulated growth factors in leiomyomas include epidermal growth factors (EGFs), fibroblast growth factors (FGFs), insulin-like growth factors (IGFs), platelet-derived growth factors (PDGFs), vascular endothelial growth factors (VEGFs), and the transforming growth factor-β family of ligands (TGF-β).

Epidermal growth factor (EGF) is a member of a family of factors that bind to four transmembrane receptor tyrosine kinases, namely epidermal growth factor receptor (EGFR/ErbB1) and human epidermal growth factor receptors 2, 3, and 4 (HER2/ErbB2, HER3/ErbB3, and HER4/ErbB4) [129,130]. EGF binds exclusively to EGFR. Transforming growth factor-α (TGF-α), amphiregulin (AREG), and epigen (EPGN), which are also members of the EGF family of ligands, also bind to EGFR/ErbB1. Other ligands that bind ErbB receptors with differing specificities include heparin-binding EGF (HB-EGF), epiregulin (EPR), betacellulin,

and neuregulins (NRGs). Cleavage of cell transmembrane-bound ligand precursors by ADAMs leads to the binding of ligands to the extracellular domain of EGF receptors. This results in receptor homo- or hetero-dimerization and the autophosphorylation of the receptor's cytoplasmic tyrosine kinase domain, leading to receptor activation. Subsequent downstream phosphorylation and activation of signalling pathways such as the Ras/MAP kinase signalling pathway, PI3K/Akt signalling pathway, phospholipase Cγ and PKC signalling pathways ensues and ultimately leads to gene transcription mediated by transcription factors belonging to AP-1, MYC, and ETS families among others [129,131].

Differential expression of EGF family members HB-EGF, EGF, and TGF-α has been demonstrated in myometrium and leiomyoma patient tissue samples, although TGF-α is shown to be expressed at low levels in myometrium and leiomyomas [132-134]. HB-EGF expression is higher in myometrium compared to leiomyomas and cells cultured from the myometrium are more responsive to HB-EGF stimulation than leiomyoma cells. In addition, EGF receptor autophosphorylation and EGF receptor-dependent leiomyoma cell proliferation is shown to be inhibited by selective EGFR tyrosine kinase inhibitors [135,136].

The fibroblast family of growth factors (FGF) is a 22-ligand family of mitogenic glycoproteins that, similarly to EGFs, bind to four transmembrane tyrosine kinase receptors, fibroblast growth factor receptors 1, 2, 3, and 4 (FGFR1-4) [137,138]. Secreted FGFs bind to heparan sulphate (HS) in the extracellular matrix and upon cleavage by heparinases, translocate to the cell surface where they form a stable FGF-FGFR-HS ternary complex. Dimerization of the complex results in the transphosphorylation of the ligand-bound FGF receptor intracellular kinase domains, which serve as phospho-specific binding sites for adaptor proteins. FGF ligand

binding to FGF receptor results in the activation of similar signalling pathways as in EGF receptor activation, namely Ras/MAP kinase signalling pathway, PI3K/Akt signalling pathway, phospholipase Cγ and PKC signalling pathways.

FGF7, 9, 12, and 13 expression is altered in leiomyomas and immunohistochemical analysis as well as immunoblot assays revealed differing FGF1 and FGF2 extracellular matrix levels in myometrium and leiomyoma samples [132,139,140]. FGF receptors 1 and 3 are also differentially expressed in leiomyomas and the inability to suppress stromal FGFR1 expression during the early luteal phase in women with leiomyoma has been linked with abnormal uterine bleeding [132,141].

Insulin-like growth factors 1 and 2 (IGF-I, IGF-II) are mitogenic ligands with high sequence similarity to insulin, and together regulate tissue-specific growth as well as organism-wide growth and energy metabolism [142]. IGF-I and II bind to two tetrameric transmembrane receptors IGF receptors (IGF1R and IGF2R) and six IGF binding proteins (IGFBP1-6), which are important in modulating receptor activity by regulating IGF levels available for binding to the receptor. Both IGF-I and II bind IGF1R whilst IGF2R prefers IGF-II, although IGF2R lacks a kinase domain and therefore acts primarily to sequester IGF-II. Downstream signalling targets include RAS/MAP kinase and PI3K/Akt/S6K kinase signalling pathways.

IGF binding proteins associate with the extracellular matrix and are proteolytically degraded by MMPs, thereby regulating the bioavailability of IGF and IGF receptor activity in smooth muscle cells and fibroblasts [143,144]. In leiomyomas, IGF1R expression was found to be increased and IGFBPs also shown to be dysregulated, with IGFBP5 found to be among the most significantly dysregulated genes [132,133,145]. Although IGF-I levels were unchanged between

fibroid tumours and matched myometrium, higher levels of IGF-I in patients were positively correlated with fibroid size [146]. In addition, IGF-II levels were increased in over two-thirds of fibroid tumours investigated.

Platelet-derived growth factors (PDGF) are disulphide-linked dimeric polypeptide ligands consisting of five proteins encoded by four genes, PDGF-A, -B, -C, and D, with PDGF-A and B capable of homo and heterodimerization [147]. The prominent and closely related vascular endothelial growth factors (VEGFs) are a subfamily of the PDGF family [148,149]. These are heparin-binding homodimeric polypeptide ligands encoded by five genes, VEGF-A, -B, -C, -D, and placenta growth factor (PlGF). Together, PDGFs and VEGFs play significant roles in vasculogenesis, angiogenesis, reproductive function and many early developmental processes. They also bind to transmembrane receptor tyrosine kinases, with PDGFs binding to PDGF-α and PDGF-β and VEGFs binding to VEGFR-1, VEGF-2, and VEGFR-3. In general, cross-family binding of receptors is not common, although exceptions have been documented, with VEGF-A demonstrated to be capable of stimulating PDGF receptors in mesenchymal stem cells [147,150]. Many of the PDGF/VEGF proteins are sequestered in the ECM and contain c-terminal basic retention motifs, CUB domains, and heparin-binding domains that are important for extracellular matrix localisation [147,151]. Extensive studies have been performed documenting the roles of PDGFs/VEGFs in tissue fibrosis [147,152,153]. Importantly, altered expression of PDGFs/VEGFs and their corresponding receptors was demonstrated by immunohistochemical and gene expression analysis in uterine leiomyomas [132,145,154,155]. Reduction in angiogenic potential of uterine fibroids in conjunction with decreased PDGF/VEGF expression was also seen in samples obtained from patients treated with a GnRH agonist demonstrating anti-angiogenic effects of GnRH therapies in

leiomyomas [156]. PDGF was also shown to be capable of stimulating VEGF receptor signalling in myometrium smooth muscle cells [157].

The transforming growth factor-β family of ligands (TGF-β) is one of the most extensively studied families of growth hormones and plays many different cellular roles including cell proliferation, differentiation, migration and death as well as varied system-wide roles such as tissue homeostasis and regeneration [158,159]. In addition, the important role performed by TGF-β growth factors in reproductive development and function is well documented [160]. Together with the PDGF/VEGF family, the TGF-β family is part of the structural superfamily of cystine knot proteins [161]. There are 33 known ligands in the TGF-β family that, based on sequence similarity and function, fall into distinct ligand subclasses: the TGF-β subfamily; Nodal and related factors; the Activin and Inhibin subfamily; the bone morphogenetic protein (BMP) subfamily and the closely related growth and differentiation factor subfamily (GDF) [162,163].

TGF-β family of ligands bind to cell surface dual-specificity receptors that possess both serine/threonine kinase activity and tyrosine kinase activity [159]. Receptors fall into two classes: seven type I receptors, known also as Activin receptor-like kinases (ALKs), and five type II receptors [164,165]. These can be further sub classified based on the type of growth factors they bind: TGF-β receptors (TβRI and TβRII) which bind TGF-β; Activin receptor I (ActR-I) binds Activins and Nodal; ActRII binds Activins, Nodal, BMP4, BMP7, GDF5 and GDF8; and BMP receptors (BMPRI and BMPRII) which bind BMPs and GDF5.

Dimeric ligand binding to the receptor is highly regulated by ligand binding proteins such as latency-associated protein (LAP), which is the cleaved pro-peptide from the TGF-β precursor

that remains associated with TGF-β and prevents its interaction with the receptor [166]. Other ligand binding proteins include Noggin, Chordin, Caronte, which bind BMPs, Follistatin that binds Activin and GDF8, and Cerberus that binds BMPs and Nodal. In addition, association of the ligand with extracellular matrix proteins such as latent TGF-β – binding proteins (LTBPs) keeps it in its inactive form and sequestered in the extracellular matrix [167]. Dissociation of the ligand from the regulatory LTBPs leads to the association of the ligand with its cognate receptor and the formation of a heterotetrameric receptor complex composed of two type I receptors and two type II receptors. This results in the phosphorylation of the type I receptor glycine-serine (GS) rich domain by the constitutively active type II receptor, an event that leads to a receptor conformational change resulting in the activation of the type I receptor and an increased affinity of the complex for the receptor substrate Smad proteins (R-Smads), which are the only known TGF-β receptor substrates.

R-Smads consist of 5 proteins: Smad2 and Smad3 which recognise TGF-β receptors, Activin, Nodal and GDF-8; Smad1, Smad5 and Smad8/9 which recognise BMP and GDF receptors [159,163,166,168-170]. The phosphorylation of cytoplasmic R-Smads is carried out by type I receptors, with the receptor regulating which R-Smad is phosphorylated in response to a specific ligand. R-Smad phosphorylation decreases their affinity for cytoplasmic retention proteins such as Smad anchor for receptor activation (SARA). R-Smad phosphorylation also exposes the Smad nuclear import sequence and increases their affinity for the common Smad, Smad4. This results in the formation and translocation into the nucleus of a heterotrimeric Smad complex composed of two R-Smads and Smad4, which then associate with cell-type specific Smad DNA-binding cofactors as well as coactivator and corepressor complexes to regulate gene transcription in a

context-dependent/ cell-specific manner [158].

Activation and nuclear translocation of R-Smads initiates a negative feedback loop involving a third group of Smad proteins, the inhibitory Smads (I-Smad). This group is composed of Smad6 and Smad7, with Smad6 primarily inhibiting BMP signalling whilst Smad7 inhibits TGF-β, GDF-8 and activin signalling. I-Smads negate TGF-β family signalling at multiple stages including preventing receptor activation through associations with type I receptors, inhibiting interaction of R-Smads with Smad4, and directly inhibiting R-Smad-dependent transcription by interacting with DNA bound R-Smad complexes [158,159]. I-Smads also recruit Smurf E3 ubiquitin ligases that mark R-Smad complexes for degradation. In addition to the primary TGF-β/Smad signalling pathway, the TGF-β superfamily is also involved in non-canonical signalling. Activation of the Erk/MAPK signalling pathway through TGF-β mediated Ras activation has been demonstrated [171]. Other non-Smad signalling pathways activated by TGF-β include the JNK/p38 signalling pathway, Rho-like GTPase signalling pathway, and PI3K/Akt signalling pathway [172,173].

TGF-β signalling has been extensively linked with multiple diseases such as inflammatory disorders, atherosclerosis, cancer, and tissue fibrosis [174,175]. In particular, the TGF-β subfamily involving TGF-β1, 2, and 3 ligands plays a prominent role in tissue fibrosis and this is primarily attributable to its regulation of extracellular matrix gene expression, with TGF-β1 highlighted as a master regulator that drives tissue fibrosis [176-178]. Through gene expression and immunoblot assays, TGF-β ligands and their corresponding receptors have been shown to be dysregulated in uterine leiomyomas [132,145,179,180]. In particular, altered TGF-β3 expression has been identified as a major cause of TGF-β – mediated fibrosis in uterine leiomyomas [181-183].

TGF-β3 treatment of myometrium and leiomyoma cells resulted in a significant increase in expression of extracellular matrix proteins that include collagens, proteoglycans, and fibronectin [184,185]. Aberrant ECM protein expression was inhibited by treatment with TGF-β3 antibody [185]. In addition, studies to determine the effects of the GnRH agonists on TGF-β signalling demonstrated the decrease in expression of TGF-β receptors and Smads upon treatment with Leuprolide Acetate [179]. TGF-β3 serum levels were also reduced in patients treated with the Ulipristal Acetate, further highlighting the therapeutic potential of targeting TGF-β signalling in uterine leiomyomas [186]. Aberrant regulation of GDF8 and Activin-mediated cell proliferation in myometrium and leiomyoma primary cells has also been reported [187]. Unlike primary cells derived from the myometrium, proliferation of leiomyoma cells was not affected by GDF8 and myostatin treatment. Activin was also shown to have a stimulatory effect on extracellular matrix gene transcription, which could be blocked by treatment with Ulipristal Acetate [187,188].

A significant number of matrisome proteins are demonstrated to be dysregulated in leiomyomas [122]. Surprisingly, despite this prominent phenotype, the mechanism by which the majority of ECM proteins are dysregulated in leiomyomas is unknown.

1.6. Leiomyoma disease genetics

High throughput sequencing techniques have provided major insights into the mechanisms of disease pathogenesis [189-191]. In uterine leiomyomas, whole exome and whole genome sequencing has catalogued multiple genetic aberrations that may be important for leiomyoma development and growth. Mehine et al. demonstrated that fibroid tumours stratify into four main subtypes that are dependent on the mutational status of fumarate hydratase (*FH*), mediator of transcription subunit 12 (*MED12*), high mobility group AT-hook 2 (*HMGA2*) translocations and collagen

(*COL4A5-COL4A6*) gene deletions [192].

Fumarate hydratase (FH) is an evolutionarily conserved homotetrameric enzyme that is required for the reversible hydration-driven conversion of fumarate to malate in the tricarboxylic acid (TCA) cycle [193]. FH is localised in the mitochondria as well as the cytosol and the presence of an N-terminal mitochondrial targeting sequence (MTS) determines the localisation of the enzyme in mammals, with differing sites of translation initiation from a single *FH* gene or differential cleavage of a single FH propeptide determining the presence or absence of the MTS [194,195]. While mitochondrial FH is involved in energy metabolism, cytosolic FH plays an important role in DNA damage response. Studies show that the induction of DNA damage response by ionising radiation-induced double strand breaks results in nuclear translocation of cytosolic FH where it is involved in DNA repair through non-homologous end-joining (NHEJ) [196]. Other studies demonstrated that the loss of cytosolic FH or impaired FH enzymatic activity and the concomitant increase in fumarate results in increased DNA double strand breaks and reduced DNA repair by homologous recombination (HR) [197,198]. In addition to its role in DNA damage response, FH is also involved in regulating the hypoxia inducible factor (HIF) pathway, with inactivation of FH shown to cause an increase in intracellular fumarate, resulting in an inhibition of hypoxia inducible factor (HIF-1α) prolyl hydroxylation [199].

Germline mutations in fumarate hydratase have been shown to predispose people to leiomyomas of the skin and uterus as well as to renal cell carcinomas [5,200,201]. *FH* mutations associated with hereditary leiomyomatosis and renal cell cancer (HLRCC) result in indels in conserved regions, loss of FH protein expression, as well as expression of a truncated protein, which all lead to a loss or severe reduction in fumarate hydratase activity. Given the role of FH

in DNA damage response, a recent study suggests that treatment of *FH*-deficient tumours with poly(ADP)-ribose polymerase (PARP) inhibitors may be therapeutically beneficial to HLRCC patients [197]. *FH* mutations were also demonstrated in groups that suffered from fumarate hydratase deficiency, which causes neurological impairment and patient life expectancy of less than a year after birth [5]. In addition to germline mutations, somatic mutations in the *FH* gene have also been identified in women diagnosed with uterine leiomyomas [202,203]. However, the mechanisms behind sporadic somatic fumarate hydratase mutations in uterine fibroids are unknown.

Chromosomal rearrangements are the structural reconstitution of chromosomes as a result of DNA double strand breaks and the aetiology of many developmental disorders and cancers has been tied to these aberrant events [204-207]. In addition, chromosomal rearrangements have been characterised in multiple benign tumours of mesenchymal origin such as lipomas, salivary gland adenomas, breast fibroadenomas as well as uterine leiomyomas [208-211]. Rearrangements can occur between chromosomes (interchromosomal rearrangement) and may be either homologous or non-homologous, or they can involve the reconstitution of sections of the same chromosome (intrachromosomal rearrangement) [212]. Intrachromosomal rearrangements are caused by one or more double strand breaks in the same chromosome and result in deletions, insertions, duplications, and inversions of parts of the chromosome. These rearrangements can occur in the interior of the chromosome (i.e. interstitial) or can occur at the ends of the chromosome (i.e. terminal), in which case the telomeres are lost.

Three types of interchromosomal rearrangements have been identified, namely i) Robertson translocations, ii) reciprocal translocations and iii) complex chromosomal

rearrangements [212]. Robertson translocations involve the breakage of chromosomes at the centromeres and reconstitution of the two long arms from the respective chromosomes to form a new large chromosome; the smaller chromosome resulting from the reconstitution of the short arms is usually lost. This form of interchromosomal rearrangement occurs only with acrocentric chromosomes. Reciprocal translocations can occur between all chromosomes and involve a single double-strand break on both non-homologous chromosomes with the detached parts being switched to opposing chromosomes upon reconstitution, which can result in gene fusions. Also, interchromosomal rearrangements of two homologous chromosomes can result in deletions or duplications.

Complex chromosomal rearrangements (CCRs) involve three or more double strand breaks on two or more chromosomes, resulting in the exchange of genetic material between chromosomes upon reconstitution [213,214]. Complex chromosomal rearrangements in foetal development are rare and occur predominantly in somatic cells and result in malignancies [215-217]. A recent study identified CCRs that result from a single event of catastrophic cellular crisis [218-220]. This singular event of chromosomal shattering, termed chromothripsis, has been identified in leiomyomas and results in chromosomal alterations at the *COL4A5–COL4A6* gene locus on the X chromosome [192]. Chromosomal shattering resulted in a *COL4A5–COL4A6* gene fusion at the 3′ ends of both genes. Also, a significant increase in transcription of the gene coding for insulin receptor substrate 4 (*IRS4*), which is located in close proximity to *COL4A5*, was observed. In addition to chromothriptic events at this locus, simple intrachromosomal deletions at *COL4A5* and *COL4A6* gene loci were observed. Similar intrachromosomal deletions of *COL4A5* and *COL4A6* have also been identified in oesophageal leiomyomas [221,222].

Another group of genes that have been identified as possible drivers of uterine leiomyoma pathogenesis and are dysregulated as a result of chromosomal shattering are the high mobility group (HMG) genes *HMGA1* and *HMGA2,* with CCRs in leiomyomas resulting in an increase in HMGA protein expression [192]. High mobility group proteins are part of a family of architectural transcription factors that include HMGA, HMGB, and HMGN families of chromatin-bound proteins [223]. They are important factors involved in multiple DNA repair pathways that include nucleotide excision repair, base excision repair, mismatch repair, and double strand break repair [224]. They also play a role in chromatin remodelling and interact with histone H1 to regulate chromatin compaction and DNA accessibility by affecting the ability of histone H1 to stabilise chromatin structure. In these ways, HMG proteins are critical components in the regulation of global gene transcription.

The HMGA family of factors possess two or three highly conserved palindromic proline-arginine-glycine-arginine-proline (PRGRP) AT-hook DNA binding motifs that preferentially bind to the minor groove of adenine-thymine (A/T) rich DNA [225]. The family is composed of four proteins, namely HMGA1a, -b, -c, which are alternative transcripts of the *HMGA1* gene, and HMGA2. They are implicated in the regulation of enhancer assembly and activity and are also shown to play a role in localised changes to chromatin structure through the recruitment of chromatin remodelling complexes [226]. HMGA2 is negatively regulated by the microRNA *let-7* and is induced by TGF-β/Smad signalling and mediates epithelial-to-mesenchymal transition in cells [227-229]. Post-translational modifications of HMGA proteins such as phosphorylation and acetylation have been shown to modulate HMGA DNA binding affinity [211].

Chromosomal rearrangements of HMGA proteins have been characterised in multiple

benign tumours as well as in some cancerous tumours [208-211,230,231]. In uterine leiomyomas, CCRs between chromosome 12 and 14 resulted in the fusion of the *HMGA2* with the DNA repair gene *RAD51B* [192]. Some rearrangements resulted in the truncation of the *HMGA2* gene, which included the excision of the *let-7b* target sequence, suggesting a possible mechanism for HMGA2 up regulation in this context. Previous work on HMGA2 expression in uterine leiomyomas suggested that *HMGA2* aberrations in uterine leiomyomas are mutually exclusive from *MED12* mutant leiomyomas [232,233]. However, recent studies demonstrate that a significant proportion of HMGA2 over expressing tumours also possess mutations in *MED12* [234-236]. Although HMGA2 up regulation is most striking in tumours possessing chromosomal rearrangements that involve HMGA2, a general increase in HMGA2 in all leiomyomas is also noted [236]. Another notable CCR occurred between chromosomes 5 and 6, resulting in the fusion of *HMGA1* with *MIR143HG,* a microRNA precursor for miR-143 and miR-145, both of which have been implicated in smooth muscle cell differentiation as well as smooth muscle maintenance and vascular homeostasis [237,238]. Despite extensive molecular characterisation of *HMGA* gene aberrations in leiomyomas, the mechanistic role of HMGA proteins in uterine fibroid pathogenesis remains largely unexplored.

In eukaryotic organisms, three RNA polymerases are responsible for transcription of DNA. RNA polymerase I (RNAPI) transcribes genes encoding ribosomal RNAs (rRNAs) for the small subunit (18S rRNA) and the large subunit (28S and 5.8S rRNAs) of the ribosome, whilst RNA polymerase III (RNAPIII) transcribes 5S large subunit rRNA, transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs). RNA polymerase II (RNAPII) is responsible for the transcription of the majority of the eukaryotic genome, being required for the transcription of messenger

RNAs (mRNAs) for all protein coding genes and pseudogenes, microRNAs, and most small nuclear RNAs [239]. RNA polymerase II is composed of DNA-directed RNA polymerase II subunits RPB1-12, encoded by the genes *POLR2A-L*. RPB1 is the largest subunit of the RNAPII complex and receives considerable attention as the main catalytic subunit of RNAPII [240]. In conjunction with RPB2, RPB1 forms the DNA binding groove of the complex as well as the magnesium ion-containing enzyme active site. Also, the C-terminal domain (CTD) of RPB1 contains multiple repeats of the consensus heptapeptide $Y^1S^2P^3T^4S^5P^6S^7$ sequence along with variations of this sequence. The presence of multiple heptapeptide repeats, of which there are 52 repeats in mammals, is unique to RNAPII and is required for transcriptional initiation and elongation as well as co-transcriptional processing of precursor messenger RNA (pre-mRNA) [241,242].

In addition to the 12-subunit RNAPII core complex, the pre-initiation complex (PIC) is also composed of general transcription factors (GTFs), which include transcription factor of RNA polymerase II, A (TFIIA), TFIIB, TFIID, TFIIE, TFIIF, and TFIIH [239,243-245]. GTFs possess DNA binding motifs that recognise core promoter sequences, with TFIID playing a primary role in this respect. TFIID recognises the downstream core promoter element (DPE), Initiator (Inr) and TATA box motifs, which are positioned downstream of the transcription start site (TSS), at the transcription start site, and upstream of the start site respectively. TFIIB recognises two TFIIB recognition elements (BREs), one that is upstream (BREu) and the other that is found downstream (BREd) of the TATA box element [245]. Sequence-specific recognition of the TATA box consensus nucleotide sequence TATAA/TAAA/G by TFIID, which is comprised of the TATA binding protein (TBP) and 13 TBP-associated factors (TAFs), is the most extensively

characterised transcription machinery promoter-binding event and is required for the assembly of the pre-initiation complex at gene promoters [246]. Precise loading of TBP to the TATA box motif is mediated by the interaction of TAFs, which bind to the downstream core promoter element and the Inr element. TBP binding to DNA forces a 90° bend of the DNA, enabling the interaction of TFIID with TFIIB, which in turn allows for TFIIB binding to the TATA-flanking BREu and BREd motifs, thereby further stabilising the budding transcription initiation complex [244]. Although not required, TFIIA also aids in stabilising the TBP-DNA interaction as well as stimulating basal and active transcription [247,248]. TFIIF forms high affinity interactions with RNAPII and aides in recruiting RNAPII to the TFIID-TFIIB-TFIIA upstream promoter complex, where the RNAPII dock domain binds to the N-terminal B-ribbon domain of TFIIB to form the core PIC [240,243]. A secondary but also important function of TFIIF is to prevent the formation of non-specific contacts between RNAPII and DNA, whilst also stabilising the PIC. Finally, TFIIE directly binds to RNAPII and recruits the multifunctional TFIIH complex to the promoter-associated core PIC.

The ten subunit TFIIH complex has three main functions in the PIC: i) to unwind DNA around the transcription start site and form a transcription bubble through its excision repair cross-complementation group 3/ xeroderma pigmentosum type B (ERCC3/XPB) subunit which has ATP-dependent 3′-5′ helicase activity, ii) to phosphorylate the C-terminal domain (CTD) of the RNAPII subunit RPB1 through the kinase activity of the MAT1–cyclin H–cyclin dependent kinase 7 (CDK7) trimeric complex, and iii) nucleotide excision repair through its ERCC2/XPD subunit that has ATP-dependent 5′-3′ helicase activity [240].

Immediately after transcription initiation, CDK7 phosphorylates Ser5 of the C-terminal domain heptapeptide repeats, leading to the recruitment of 5′ pre-mRNA capping enzymes to the nascent transcript exiting the RNAPII exit channel [241,249-251]. CDK7 has also been shown to phosphorylate Ser7 of the C-terminal domain heptapeptide repeats in a gene specific manner, which may play a role in the transcription of snRNAs [241]. Interestingly, at this stage of transcription, instead of entering a phase of transcriptional elongation of the nascent transcript, a stably associated DNA–PIC–mRNA complex remains paused at the gene promoter, an event that is shown in part to be important in ensuring complete mRNA capping before entering elongation [252]. This promoter proximal pausing is initiated and maintained by negative elongation factor (NELF) and DRB (5,6-dichloro-1-β-d-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF) complexes, which together bind to RNAPII and prevent elongation [253]. The paused state is maintained until positive transcription elongation factor-b (P-TEFb), which is composed of cyclin T1 and cyclin dependent kinase 9 (CDK9), phosphorylates NELF and Ser2 of the RNAPII C-terminal heptapeptide repeats, thereby evicting NELF from the promoter and releasing RNAPII from the paused state [253,254]. In addition, P-TEFb also phosphorylates DSIF, converting it into a positive elongation factor. This series of events allows transcriptional elongation to proceed to termination, after which RNAPII is recycled for a new round of transcriptional initiation.

Despite the necessity of RNA polymerase II and the general transcription factors, the pre-initiation complex on its own is insufficient to stimulate eukaryotic gene transcription. An additional complex, mediator of RNA polymerase II transcription (Mediator), is required for competent gene transcription [255]. Core Mediator is a 26-subunit complex that functions as a

bridge between transcription factors, cofactors and the pre-initiation complex, thereby relaying context-specific and calibrated signals from promoter-distal regulatory sites in the genome to the promoters of active genes. It is organised into three functional modules, which are designated as the Head, Middle and Tail modules. The Head and Middle modules make direct contacts with RNAPII during initiation, binding to a hypo-phosphorylated RNAPII and stimulating RNAPII C-terminal domain phosphorylation by TFIIH, which then results in the release of Mediator from the PIC [256,257]. The tail module interacts with transcription factors and cofactors, thereby stimulating transcription in a transcription factor-specific manner. Whilst the structure of the individual subunits within the modules is fairly stable, the three modules undergo conformational changes relative to each other that influence the interactions of Mediator with other complexes.

In addition to the core Mediator Head, Middle and Tail transcription-activating modules, a fourth module exists that has primarily transcription-repressive functions. The CDK8 subcomplex is a four subunit module comprising of CDK8, cyclin C, MED12 and MED13, with the two Mediator subunits acting as a bridge between the core Mediator complex and the CDK8 subcomplex [255,257,258]. Although core Mediator interacts with RNA polymerase II through the Head/Middle modules and with the CDK8 submodule through the Tail module, studies have shown that core Mediator complex interaction with RNA polymerase II and the CDK8 submodule are mutually exclusive [259]. Thus, the promoter-proximal Mediator complex exists in two forms, a gene activating core Mediator complex that is able to engage RNA polymerase II and a RNA polymerase II-refractory Mediator-CDK8 complex.

Recently, the mediator complex subunit 12 gene (*MED12*) was found to be highly mutated in leiomyomas [260]. In 80 patient samples consisting of 225 fibroids, 75% of all patients

had tumours that had a mutation in *MED12*. Interestingly, most of these mutations were located in exon 2 of *MED12* (70%), with a glycine mutation at codon 44 (G44) being the most frequently mutated at 49%. Multiple follow up studies have confirmed this finding in uterine leiomyomas in patient populations around the world [261-263]. In addition to uterine leiomyomas, similar mutations in *MED12* have been characterised in breast fibroadenomas, leiomyosarcomas, and colorectal cancer [264,265]. Mutations in *MED12* have also been identified in prostate cancers and multiple developmental disorders, albeit in different locations on the *MED12* gene [266-271]. Further underlining the role of *MED12* mutations in leiomyomas, a recent study demonstrated that introduction of a point mutation that causes a glycine to aspartate change in codon 44 (G44D) is sufficient to cause leiomyomas in mice [272]. While loss of wild type *Med12* gene expression did not result in leiomyoma formation in mice, expression of a G44D mutant Med12 protein in mice with a *Med12* wild type knockout background resulted in the development of leiomyomas.

Also, previously published interaction studies in heterologous cell lines expressing mutant MED12 (G44D/S) have suggested that mutations in exon 2 of MED12 may result in a decrease in CDK8 submodule binding affinity at sites of active transcription [273]. Given that MED12 exon 2 forms the cyclin C binding pocket, affinity purification mass spectrometry in heterologous cell lines expressing mutant MED12 demonstrated that mutations in exon 2 of MED12 results in a decrease in binding affinity between cyclin C and MED12. However, recent studies have shown the ability of MED13 to ameliorate the cyclin C-MED12 G44 mutation-dependent binding defect and stabilise the CDK8 submodule [274]. Despite knowledge of the role of MED12 in transcriptional regulation and evidence of MED12 mutations in leiomyomas, there

still exists a gap in knowledge with regard to how MED12 mutations perturb gene regulation pathways, thereby leading to uterine leiomyomas.

1.7. Epigenetic regulation of uterine leiomyomas

Multiple synergistic mechanisms of transcriptional regulation in cells have been characterised, all of which play an important role in the finely tuned regulation of gene expression. Epigenetic regulation, which is broadly defined as the "mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" [275], is of increasingly great interest as an important mechanism by which gene expression is regulated in cells [276-278].

DNA methylation, a heritable covalent modification of DNA and an important mechanism of gene silencing, is an archetypal example of epigenetic gene regulation. Methylation of the cytosine pyrimidine ring at the C5 position (5mC) is the predominant form of DNA methylation in eukaryotes, although methylation of the adenine amino group at the C6 position (6mA) has also recently been identified [279-282]. While DNA methylation is generally considered to be refractory to gene transcription, studies show that its impact is context specific. DNA methylation at the transcription start site prevents transcription initiation while methylation of the gene body does not seem to negatively impact transcription and is a fairly common feature in transcribed genes [283]. DNA methylation of gene bodies has also been shown to increase the rate of cytosine-to-thymine conversions in somatic cells [284]. Many gene promoters contain short 1-2kb cytosine-guanine dinucleotide (CpG) regions of DNA, the majority of which remain unmethylated. Such CpG islands (CGIs) form an important mechanism of long-term gene silencing as seen in random X inactivation, cell type-specific gene silencing, and genomic imprinting. CGIs are also present in the gene body with lower frequency and although their

function in this context is not fully understood, it has been suggested that they play a role in gene splicing and are also important in the repression of transposable elements, thereby aiding in maintaining genome stability [283]. *De novo* DNA methylation, as well as maintenance DNA methylation, are established by DNA nucleotide methyltransferases (DNMTs) such as DNMT1, -2, -3a, -3b, and -3L, which are responsible for 5mC deposition, as well as methyltransferase like 4 (METTL4) and N6 adenine-specific DNA methyltransferase 1 (N6AMT1), which are responsible for 6mA deposition [285-287]. In addition, DNA demethylases and DNA methylation-specific binding proteins have been identified, together forming a complex system for the deposition, interpretation and removal of methyl marks on DNA [280,288,289].

In uterine leiomyomas, an altered DNA methylation profile has been shown relative to normal matched myometrium samples [290,291]. Hypermethylated promoters accounted for 84 out of 120 differentially methylated promoters whose gene expression was also significantly altered, with collagen genes appearing as notable inclusions in this list. Oestrogen receptor target genes were also prominent among gene promoters with differential DNA methylation. Although DNA hypermethylation of gene promoters is more common in leiomyomas, DNA hypomethylation has also been shown [290,292]. Interestingly, the X chromosome has been identified as a hotspot for DNA hypomethylation in leiomyomas, although the significance of this is unclear as the expression of hypomethylated genes was unchanged and imprinted genes on the X chromosome were unaffected. Studies have also identified aberrant expression of DNA methyltransferases and demethylases in uterine leiomyomas, providing a possible mechanism for the alterations in methylation profiles between myometrium and leiomyomas [293]. Despite a tremendous amount of focus on DNA methylation in uterine leiomyomas, only a handful of dysregulated genes are

identified with hypo/hypermethylated DNA in fibroids. In addition, the relationship of these genes to leiomyoma disease pathogenesis remains unclear.

In eukaryotic organisms, DNA is a linear, organised structure that is tightly bound by proteins and packaged to form chromatin [294,295]. This high level of organisation allows for the efficient packaging of DNA into the nucleus as well as enabling a higher level of gene regulation whilst protecting DNA from damage. The main proteins associated with DNA in chromatin structure are histones. They are basic molecules, which enables them to interact with the negatively charged DNA. Five main histone variants exist, namely histone H1, H2A, H2B, H3 and H4. Histones H2A, H2B, H3 and H4 are known as the core histones and two molecules of each core histone combine to form a histone octamer around which 147 base pairs of DNA wrap to form a nucleosome, the basic unit of chromatin fibre [276]. Histone H1 binds to linker DNA that exists between nucleosomes and has an approximate length of 20-80 base pairs. Histone H1-bound linker DNA, together with the core nucleosome, forms an 11nm primary nucleosome structure [296,297]. Multiple nucleosomes wrap into a 30nm fibre, which constitutes the secondary structure of chromatin that is further folded into a compact higher order tertiary structure. The level of chromatin secondary and tertiary structure compaction regulates the accessibility of DNA and the most compact tertiary structure is referred to as heterochromatin. This arrangement is associated with a repressive gene state, unlike DNA-accessible euchromatin that contains inducible genes, although recent evidence details actively transcribed genes in heterochromatic regions as well [298,299].

Histone proteins possess a globular core domain as well as an unstructured N-terminal tail. The globular domain is important in mediating histone-histone interactions as well as

histone-DNA interactions. The unstructured histone tails have been extensively studied and shown to play an important role in regulating gene activation and repression [300]. The location of the unstructured histone tails outside of the 30nm fibre leaves the tails readily accessible for modification, although modifications of globular domain residues has also been documented [301]. Posttranslational modifications (PTMs) of histone tails are important for regulating transcription, DNA damage repair, as well as chromosome condensation and replication [302]. In transcription, posttranslational modification of histone tails alters the level of chromatin compaction, thereby controlling the accessibility of DNA to transcription factors, chromatin remodelling complexes and RNA polymerase II.

Histones are modified in response to cellular cues that result in transcription factors translocating into the nucleus and binding to DNA and in the process also recruiting coactivating or corepressing chromatin-modifying complexes. The complexes then alter the nucleosome surface through the enzymatic addition of PTMs on histone tails, which are then recognised by chromatin interacting proteins in a modification-specific manner. This results in the recruitment of transcription regulating complexes as well as the recruitment of additional chromatin modifying complexes such as chromatin-remodelling enzymes that further alter the nucleosome surface. ATP-dependent chromatin-remodelling enzymes modify nucleosome density in a region by evicting nucleosomes or by altering the spacing between nucleosomes, thereby creating or destroying nucleosome free regions (NFRs) along the DNA on which transcription factors can bind. The creation of nucleosome free regions is key for efficient gene transcription as nucleosomes are in general refractory to transcription, either through physically obstructing factor binding to DNA or bending DNA into a conformation that impedes factor binding.

Many histone posttranslational modifications have been identified and include prominent modifications such as methylation, acetylation, and phosphorylation [300]. Less characterised but also important PTMs include ubiquitination, SUMOylation, and ribosylation [302]. Even less known and less characterised modifications have been identified and work to describe their role in the cell continues [303,304]. Histone PTMs can be broadly categorised into gene repressive and gene activating marks and, in many cases, different modifications of the same residue fall into opposing categories. Lysine residues are an archetypal example of histone posttranslational modification. They are the most frequently modified and are also capable of undergoing the most diverse range of posttranslational modifications, with each modification playing a potentially different role in a context dependent manner. Methylation and acetylation of histone lysine residues are extensively studied PTMs, with acetylation of lysine generally considered an active mark of transcription whilst lysine methylation can be both repressive and activating, depending on the lysine residue. For example, methylation of lysine 9 and 27 on histone H3 and lysine 20 on histone H4 are known to be repressive marks, whereas methylation of lysine 4, 36 and 79 on histone H3 are all marks associated with active gene transcription [305,306]. All histone posttranslational modifications are reversible. As such, enzymes responsible for the deposition of the PTMs as well enzymes responsible for their removal have been identified [300,302].

Cis-regulatory DNA elements (CREs) are regions of DNA that contain transcription factor recognition sequences (motifs) that allow sequence-specific DNA binding by the factors [307,308]. Gene promoters, which contain binding motifs for general transcription factors, constitute the most well characterised *cis*-regulatory elements. In addition to promoters, CREs can be located in the gene body (intragenic) or they can be found in genomic regions between genes

(intergenic). As only a small proportion of the genome is made up of protein coding regions, the majority of CREs are located in non-coding intergenic regions, with gene deserts, which are very large areas of chromosomes that contain relatively few genes, also identified as possible reservoirs for CREs [309]. They are also either in fairly close proximity to gene promoters and the TSS (promoter-proximal) or located at remote distances from promoters (promoter-distal). Importantly, CREs are involved in gene activation (enhancers), gene repression (silencers), or the structural organisation and maintenance of intrachromosomal regulatory domains (insulators) [308,310,311]. Enhancers and silencers are structurally very similar and differ only in the transcription factor motifs they contain and the corresponding gene activating or repressing factors that bind to them. However, enhancers have received considerably more attention and recent studies have demonstrated that the majority are species specific, despite the similarity in the genes they regulate as well as the similarity in the expression patterns of those genes [312-314]. Underlying the importance of enhancers in eukaryotic gene regulation are studies that demonstrate that the deletion of ultra-conserved enhancers results in developmental abnormalities [315]. In addition, mutations in ultra-conserved enhancers have been shown to be responsible for evolutionary traits such as limb loss in a subset of reptilian species [316].

Gene transcription is very signal specific, with different cellular cues resulting in similar or opposing effects to the same gene in a spatiotemporally precise manner. This high level of control is mediated by ~1,600 transcription factors that bind to a variety of enhancers [317]. Given that there are ~43,000 protein coding and non-coding genes in the human genome that are expressed at different levels and for differing lengths of time in response to cellular cues, most genes are found to be associated with multiple enhancers that have a variety of transcription

factor binding motifs in diverse combinations working in concert to enable a finely tuned transcriptional output. Similarly, some enhancers associate with more than one gene, thereby serving as a hub from which the same signalling cues can be relayed to multiple genes at the same time. Through the numerous permutations of possible enhancers regulating transcription, cell type-specific gene expression is achieved [318]. Some enhancers result in the same transcriptional output for specific genes and this provides redundancy protection by relaying very similar cues to gene promoters. Enhancers usually have multiple transcription factor binding motifs for different factors clustered close together and can also have repeats of the same motif, which facilitates increased binding of the same transcription factor at the enhancer.

Genome-wide analysis has revealed the presence of histone posttranslational modifications at gene promoters and they have been demonstrated to be integral for the recruitment of complexes that are involved in transcription initiation and elongation [300,319]. Additionally, the presence of histone PTMs at promoter-distal CREs has also been catalogued [320]. Different histone PTM signatures exist for promoters and enhancers, which vary depending on gene activation or repression. Active promoters and enhancers both have acetylated lysine at position 27 of the histone H3 tail (H3K27Ac) and methylation at lysine 4 (H3K4me), but promoters undergo tri-methylation (H3K4me3) whilst enhancers are only mono-methylated (H3K4me1) [319,321,322]. While the H3K4 methylation at gene promoters has been shown to be important for competent gene transcription, H3K4 methylation at enhancers is dispensable, with the presence of the methyltransferase, functioning as a coactivator independent of its enzymatic activity, being sufficient for transcription of enhancer gene targets [323]. In addition to H3K4me3, many other histone PTMs have been identified at active promoters [300,319].

In the nucleus, enhancers and silencers spatially organise in close proximity to their respective promoters and this enables direct contact between trans-acting factors and gene promoters. This organisation is achieved through the looping of chromatin and is a general mechanism by which transcription factors bound at distal sites can influence transcription at gene promoters [324-326]. Recent work shows that sustained and stable interaction of enhancers with their corresponding promoters is required for productive transcription [327]. Intrachromosomal interactions, which are chromatin interactions that occur within a chromosome, are the predominant type of contacts formed between regions of chromatin, with relatively few interchromosomal interactions being identified in cells. In addition to chromatin interactions that occur between promoters and their enhancers or silencers (E-P or E-S), interactions between promoters (P-P) as well as interactions between enhancers (E-E) also occur. Promoter-promoter interactions between transcribed genes are common in the genome and inter-gene regulation through these interactions has been described [328]. Although enhancer-enhancer interactions are also very common and are predicted to play a role in gene regulation, more work is required to characterise their function in transcription [329].

While enhancer-promoter contacts occur over very large distances, a constraint is placed on the maximum distance of interactions, with an upper limit being 1-2Mb. This is a result of architectural proteins such as CCCTC-binding factor (CTCF) and Cohesin, which bind to insulator *cis*-regulatory elements and form long-range CTCF-CTCF intrachromosomal contacts [330]. This creates insulated neighbourhoods known as topologically associated domains (TADs) in which CREs within a domain are insulated from contacts with other CREs that are outside of the domain. Furthermore, evidence detailing the transcriptional co-regulation of genes suggests that

TADs are not only structural boundaries within chromosomes but also a means of synergistic transcriptional control of genes within the constrained 1-2Mb region of DNA [328,331].

In addition to occupying regions that establish boundaries for topologically associated domains, CTCF binding at enhancers has also been shown, although significantly less CTCF is bound at intradomain sites than at insulators [332-334]. Ren et al. demonstrated the intra-TAD binding of CTCF stabilises enhancer-promoter contacts thereby enabling consistent gene transcription at promoters [333]. The importance of CTCF in this role is underscored by an increase in cell-to-cell variability in gene expression upon the loss of CTCF binding at enhancer sites. In addition to CTCF, chromatin immunoprecipitation of H3K27Ac and H3K4me3 followed by mass spectrometry (ChIP-MS) identified the transcription factor Yin Yang 1 (YY1) as another protein that is required for the formation of enhancer-promoter contacts [334]. YY1, a multifunctional protein, is ubiquitously expressed in all cell types and is bound at distal enhancers and promoter proximal CREs and facilitates chromatin looping between enhancers and promoters. The depletion of YY1 or deletion of YY1 motifs at gene promoters results in a loss of chromatin looping ability and demonstrates the importance of YY1 as a general regulator of enhancer-promoter loops.

Aberrant epigenetic modification deposition and transcription factor chromatin binding have long been implicated in disease development [335]. In addition to epigenetic modifications at promoters such as DNA methylation, changes in chromatin structure at enhancer regions have recently been demonstrated to play a major role in disease pathogenesis [336]. Despite this, the role of enhancer chromatin structure in uterine leiomyomas remains entirely unexplored.

1.8. Activator protein-1 in leiomyoma disease pathogenesis

The activator protein-1 (AP-1) complex is a transcription factor composed of homodimers or heterodimers from the JUN, FOS, activating transcription factor (ATF), and musculoaponeurotic fibrosarcoma (MAF) families of transcription factors [337-340]. AP-1 is responsive to varied stimuli and is important in cell proliferation, differentiation, oncogenesis, survival, and apoptosis [341]. AP-1 subunits possess a basic leucine zipper (bZIP) domain which is required for dimerization of subunits and for AP-1 recognition of three very similar palindromic DNA binding motifs: the cAMP response element (CRE, 5′-TGACGTCA-3′), the 12-O-tetradecanoylphorbol-13-acetate (TPA) response elements (TRE, 5′-TGAG/CTCA-3′) and MAF recognition elements (MAREs), which can be further subdivided based on whether they contain CRE (C-MARE, 5′-TGCTGACGTCAGCA-3′) or TRE (5′-TGCTGAG/CTCAGCA-3′) core sequences [340,342]. Different combinations of the AP-1 dimer have differing dimeric stabilities as well as varied DNA specificity, and this also dictates their role in the cell. While JUN family members are capable of forming stable homodimers, FOS family members are widely thought to be unstable and therefore incapable of homodimerization and function primarily by heterodimerization with JUN proteins. However, recent evidence suggests that some FOS family members are capable of homodimerization, albeit with higher dimer dissociation constants [343]. Live cell imaging of FOS demonstrated its ability to form stable homodimers that are capable of interacting with chromatin.

The dimeric composition of AP-1 determines which AP-1 responsive genes are regulated and whether the resulting AP-1 complex is gene activating or repressive. JUN, FOS, and FOSB are generally considered to have strong transactivation properties whilst JUNB, JUND, FRA-1

and -2 are weakly transactivating or trans repressive in a signal dependent manner [341]. AP-1 activity is modulated through transcriptional regulation of AP-1 subunit gene expression and post-transcriptionally by regulation of AP-1 subunit mRNA stability. In addition, post-translational regulation of AP-1 is achieved through phosphorylation of subunit transactivation domains by mitogen-activated protein kinases (MAPKs) such as JUN N-terminal kinases (JNKs) and extracellular signal regulated kinases (ERKs).

AP-1 is activated by varied stimuli that include growth factors, cytokines, hormones, and by bacterial or viral infections. It interacts with many transcription factors and thereby mediates signal dependent transcriptional output of many signalling pathways. In particular, AP-1 has been demonstrated to form complexes with SMAD family proteins, nuclear receptors such as glucocorticoid receptor (GR), oestrogen receptor (ER), and the retinoic-acid receptor (RAR and RXR), and together carry out transactivating functions in cells [344-346]. Importantly, AP-1 subunits have been shown bind to a cAMP response element on the MyoD promoter and modulate MyoD expression in muscle cells [347]. Reciprocally, MyoD also modulates expression of AP-1 subunits [348,349]. The AP-1 complex also directly interacts with MyoD and is important for muscle cell differentiation [344,350]. Overall, these studies demonstrate an important role for AP-1 in muscle cell development.

In addition to the defined role played by AP-1 in muscle cell development, AP-1 has previously been implicated in the development of many fibrotic diseases [351-354]. Importantly, down regulation of JUN and FOS mRNA levels has been observed in uterine fibroids [355-357]. However, the mechanisms by which AP-1 down regulation lead to changes in uterine leiomyoma gene transcription are not clearly understood.

1.9. Transcriptomic and epigenomic profiling of uterine leiomyomas

In addition to whole genome and whole exome sequencing, gene expression profiling of leiomyomas has also been performed, primarily using RNA microarrays [132,145,192,355,356,358,359]. Recent work has revealed subtype-specific gene expression profiles, indicating possibly different mechanisms of pathogenesis in fibroid tumours belonging to different subtypes [132]. Despite multiple gene expression studies identifying aberrant gene regulation, very little is known regarding the mechanisms of gene dysregulation in uterine leiomyomas.

Recent chromosome conformation capture techniques have led to high resolution maps of enhancer-promoter contacts, allowing for the unambiguous pairing of enhancers with their corresponding promoters [360-363]. In addition, promoter capture Hi-C (CHi-C) has been used to demonstrate the dynamism of enhancer-promoter contacts during cell differentiation [364,365]. However, the existence of dynamic or altered contacts in human disease as a result of epigenetic changes to chromatin has not been investigated.

In this study, freshly procured tissue samples from women who had undergone hysterectomies as a course of treatment for uterine leiomyomas were used to characterise epigenetic changes in the disease. As the majority of women diagnosed with uterine fibroids have tumours with a mutation in *MED12* and the predominant mutations result in an amino acid substitution at glycine 44 (G44) in exon 2, fibroid tumours confirmed to have a glycine-to-aspartate (G44D) or glycine-to-serine (G44S) substitution in *MED12* were taken as representative tissue samples for the disease. Adjacent, non-diseased areas of the myometrium from the same patients were also collected to represent normal (WT) samples.

Previous studies suggest that culturing of primary cells from tissue samples may partially

alter the gene expression profiles of cells [366]. In primary cell culture of *MED12* mutant cells

obtained from leiomyoma patient tissue samples, a rapid loss of *MED12* mutant cells was

observed, suggesting a very limited viability of this cell population in culture [367,368]. In an effort

to avoid artifactual alterations to the transcriptomic and epigenomic profiles of patient samples,

gene expression profiling by RNA-sequencing and epigenetic profiling by high-resolution ChIP-

sequencing were performed directly from tissue samples with minimal processing. In addition, a

modified protocol optimised for work with tissue samples was used to perform promoter capture

Hi-C directly from patient tissue, allowing for the native characterisation of intradomain

chromatin architecture changes in uterine leiomyomas. These methods should be applicable to all

solid tumours and provide a reproducible approach for the characterisation of solid tumours

without the need for primary cell culture.

CHAPTER 2. MATERIALS AND METHODS

2.1. Collection of uterine tissue samples

Tumour samples were collected from 15 premenopausal women undergoing hysterectomies as a treatment for uterine fibroids. Normal, adjacent myometrium tissue samples were also collected from hysterectomies. Sanger sequencing of PCR amplified genomic DNA extracted from leiomyomas and matched myometrium was used to positively identify *MED12* mutant leiomyoma samples with either a G44D or G44S mutation. Tissue samples were either immediately cryopreserved at -80°C, processed for high-resolution ChIP-seq, RNA-seq and promoter capture Hi-C, or processed for primary cell culture. Tissue from all 15 patients was used in RNA-sequencing experiments and tissue from 5 of the patients was used for ChIP-seq and capture Hi-C experiments. All surgeries were performed at Northwestern University Prentice Women's Hospital and all patients gave informed consent for their participation in this study. The study was carried out in accordance with a Northwestern University institutional review board approved protocol. Samples were collected from hysterectomies performed on 8 black women, 3 white women, and one latino woman. The race/ ethnicities of 3 women were unknown. Patient ages ranged from 38 – 51 years with a mean of 46.1 years.

2.2. Tissue digestion and primary cell culture

Myometrium and leiomyoma tissue were cut into 5mm^3 pieces and digested with shaking (100rpm) at 37°C for 5 hours in collagenase digestion buffer (Phenol red-free Hanks balanced salt solution [ThermoFisher, cat # 14025], 2500KU deoxyribonuclease I [Sigma-Aldrich, cat # D5025], 1.5mg/ml collagenase [Sigma-Aldrich, cat # C0130], 3mM CaCl$_2$) at a ratio of 1 part w/v tissue with 5 parts digestion buffer. Digested tissue was filtered with sterile gauze sponges and centrifuged for 5min at 400 x *g*. Cells were resuspended in smooth muscle cell culture media

(SmGM-2 smooth muscle cell growth medium BulletKit [Lonza, cat # CC-3182], 1% Penicillin-Streptomycin), filtered with a 70μm cell strainer, plated and grown at 37°C in a humidified cell culture incubator containing 5% CO_2. Media was changed every 48 hours and cells were passaged at 90% confluence. Primary cells were used in experiments after two passages.

2.3. Tissue RNA extraction and sequencing

30mg of myometrium and leiomyoma tissue was cryopulverised into a fine powder using a Covaris cryoPREP CP02 impactor and homogenised for 1 minute at power setting 22 in Qiazol lysis reagent using a Kinematica Polytron PT2100 homogeniser. Tissue sample was further homogenised and supernatant isolated by room temperature centrifugation at 20,000 x g using a Qiagen QIAshredder (Qiagen, cat # 79656). RNA was isolated using Qiagen miRNeasy Mini kit according to manufacturer's instructions (Qiagen, cat # 217004). After RNA elution, an in-solution DNase treatment was performed to completely digest genomic DNA (Qiagen, cat # 79254). RNA quality was verified using the Bioanalyzer eukaryote total RNA 600 nano kit (Agilent technologies, cat # 5067). Purified RNA samples had a mean RNA integrity number (RIN) of 8.0. RNA was prepared for sequencing using the KAPA Stranded RNA-seq with RiboErase kit according to the manufacturer's instructions (Kapa Biosystems, cat # KK8483). Paired-end sequencing of all RNA libraries was performed on the Illumina NextSeq 500 platform.

2.4. Tissue preparation for chromatin immunoprecipitation and capture Hi-C

Myometrium and leiomyoma tissue samples were cryopulverised into a fine powder using a Covaris cryoPREP CP02 impactor followed by a mortar and pestle. Samples were immediately formaldehyde cross-linked (1% formaldehyde in phosphate buffered saline) for 10min at room

temperature with rotation and then quenched with 0.15M glycine for 10min at room temperature with rotation. Cross-linked tissue was recovered by centrifugation at 1,000 x *g* at 4°C for 5min and the supernatant discarded. Two washes in phosphate buffered saline containing 1X Protease inhibitor cocktail (Roche, cat # 4693132001) was performed and tissue recovered by centrifugation at 1,000 x *g* at 4°C for 5min. Cross-linked tissue was then either frozen at -80°C or immediately used in ChIP or CHi-C experiments.

2.5. Tissue chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitation was adapted from previously described methods [369]. Cross-linked myometrium and leiomyoma tissue samples were dounce homogenised in ChIP lysis buffer 1 (50mM HEPES-KOH [pH 7.6], 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% IGEPAL CA-630, 0.25% Triton X-100, 1X Protease inhibitor cocktail [Roche, cat # 4693132001]) followed by end-over-end rotation at 4°C for 15min and then centrifugation for 5min at 1,000 x *g*, also at 4°C. Samples were resuspended in ChIP lysis buffer 2 (10mM Tris-HCl [pH 8.0], 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 1X Protease inhibitor cocktail [Roche, cat # 4693132001]) followed again by end-over-end rotation at 4°C for 15min and sample recovery by centrifugation at 1,000 x *g* at 4°C for 5min. Samples were then resuspended in MNase digestion buffer (20mM Tris-HCl [pH 8.0], 150mM NaCl, 2mM EDTA, 1% Triton X-100, 0.1% SDS, 125U MNase [Worthington, cat # LS004798], 1X Protease inhibitor cocktail [Roche, cat # 4693132001]) and incubated at 37°C until 75% mononucleosomal chromatin profile is observable by gel electrophoresis of purified MNase digested DNA (approximately 5-7min). Digestion was quenched with MNase quenching buffer (10mM EDTA, 20mM EGTA) after which samples were briefly sonicated in an ice water bath (Misonix, setting 6 [~6 W power

output], 3 cycles of 15sec on and 45sec off). Collagen and cell debris were removed by centrifugation at 20,000 x g at 4°C for 20min, with recovery of chromatin-containing supernatant. Solubilised chromatin concentration was measured by BCA assay according to manufacturer's instructions (ThermoFisher, cat # 23225). Approximately 400μg of chromatin and 3μg of antibody against H3K27Ac (Active motif, cat # 39685) was used for overnight immunoprecipation of histones at 4°C with end-over-end rotation. 600μg of chromatin and 5μg of antibody was used for overnight immunoprecipation of factors. Antibodies against RNAPII, FOS (Millipore, cat #s 05-952, 06-341 respectively), JUN (abcam, cat # ab31419), CDK8 and MED12 (Bethyl, cat #s A302-500A, A300-774A respectively) were used in pulldowns. 10μl of protein G magnetic beads (ThermoFisher, cat # 10004D) per μg of antibody was added per sample and incubated at 4°C for an additional 3 hours. Immunoprecipitated samples bound to beads were recovered using a magnetic rack (ThermoFisher, cat # 12321D), followed by 5 washes with ChIP-RIPA wash buffer (50mM HEPES-KOH [pH 7.6], 500mM LiCl, 1mM EDTA, 1% IGEPAL CA-630, 0.7% sodium deoxycholate) and once with NaCl containing TE buffer (10mM Tris-HCl [pH 8.0], 1mM EDTA, 50mM NaCl), with magnetic recovery of immunoprecipitated samples bound to beads after each wash. DNA was eluted off the beads twice with 50μl ChIP elution buffer (0.1M NaHCO$_3$, 1% SDS) at 65°C for 20min with shaking. Cross-links were reversed with 300mM NaCl for 12-16 hours at 65°C followed by 20μg of RNase A (Worthington, cat # LS002132) treatment at 37°C for 1 hour. This was followed by 2 hours at 55°C with 80μg of proteinase k (ThermoFisher, cat # 25530015) supplemented with 16.5mM EDTA and 66mM Tris-HCl (pH 8.0). Reverse cross-linked DNA was isolated by phenol-chloroform extraction with phenol/chloroform/isoamyl alcohol [25:24:1] (Sigma cat, #

77617) using 5Prime phase lock gel (Quantabio, cat # 2302830) according to manufacturer's instructions, followed by ethanol precipitation (2X volume of 100% ethanol, 0.1X volume of 3M sodium acetate, and 2μl of glycogen) at -80°C for 1 hour. DNA was then pelleted by centrifugation at 20,000 x *g* at 4°C for 20min, followed by one wash in with 70% ethanol and further centrifugation at 20,000 x *g* at 4°C for 10min. DNA pellet was dried and resolubilised in 50μl of elution buffer (10mM Tris-HCl, pH 8.0 - 8.5). KAPA Hyper Prep kit (Kapa Biosystems, cat # KK8502) was used for end-repair, A-tailing, and adapter ligation with TruSeq index adapters, all according to manufacturer's instructions. A 0.6X-1.0X AMPure XP bead (Beckman Coulter, cat # A63881) size selection was carried out after adapter ligation to enrich for mono-nucleosomal DNA fragments from the H3K27Ac ChIP followed by 10-12 cycles of PCR amplification. To enrich for sub-nucleosomal DNA for use in high-resolution mapping of transcription factors and cofactors, a 1.1X-1.8X AMPure XP bead (Beckman Coulter, cat # A63881) size selection was carried out after adapter ligation followed by 10-12 cycles of PCR amplification. Single-end sequencing of all ChIP libraries was performed on the Illumina NextSeq 500 platform.

2.6. Promoter capture Hi-C

Promoter capture Hi-C was adapted from previously described methods [363,370-373]. Cross-linked myometrium and leiomyoma tissue samples were dounce homogenised in permeabilisation buffer (10mM Tris-HCl [pH 8.0], 10mM NaCl, 0.2% IGEPAL CA630, 1X Protease Inhibitor cocktail) followed by end-over-end rotation at 4°C for 15min after which sample pellet was recovered by centrifugation at 4°C. Samples were washed in 1.2X NEBuffer2 (cat # B7002) at 4°C for 15min with end-over-end rotation followed by centrifugation at 4°C. Samples were then

resuspended in 1.2X NEBuffer2 containing 0.2% SDS buffer and incubated at 37°C for 1 hour

with end-over-end rotation followed by the addition of 1.8% triton X-100 at 37°C for an

additional hour. The resulting chromatin was digested with 400U of *Hin*dIII restriction enzyme

(NEB cat # R0104) for 12-16hrs at 37°C with end-over-end rotation. This was followed by

DNA polymerase I–mediated (50U [NEB, cat # M0210]) labelling of DNA ends with dNTPs

(28.4μM D-Desthiobiotin-7-dATP [Jenna Bioscience, cat # NU-835-Desthiobio], 28.4μM dCTP

[Invitrogen, cat # 18253-013], 28.4μM dGTP [Invitrogen, cat # 18254-011], 28.4μM dTTP

[Invitrogen, cat # 18255-018]) for 3 hours at 37°C. Samples were then incubated at room

temperature for 4hrs with end-over-end rotation in proximity ligation mix (1X T4 DNA Ligase

buffer [NEB, cat # B0202], 1X BSA [NEB, cat # B9001], 2000U of T4 DNA Ligase [NEB, cat #

M0202]) followed by the addition of SDS to a final concentration of 1%. Samples were then

treated with 50μg of RNase A (Worthington, cat # LS002132) at 37°C for 1 hour with end-over-

end rotation followed by treatment with 1mg of proteinase k (ThermoFisher, cat # 25530015) at

55°C for 2 hours with shaking. Cross links were reversed with 400mM NaCl for 12-16 hours at

65°C. Labelled and proximity ligated DNA was isolated by phenol-chloroform extraction with

phenol/chloroform/isoamyl alcohol [25:24:1] (Sigma cat, # 77617) using 5Prime phase lock gel

(Quantabio, cat # 2302830) according to manufacturer's instructions, followed by ethanol

precipitation (2X volume of 100% ethanol, 0.1X volume of 3M sodium acetate, and 2μl of

glycogen) at -80°C for 1 hour. DNA was then pelleted by centrifugation at 20,000 x *g* at 4°C for

20min, followed by one wash with 70% ethanol and further centrifugation at 20,000 x *g* at 4°C

for 10min. DNA pellet was dried and resolubilised in 200μl of elution buffer (10mM Tris-HCl,

pH 8.0 - 8.5). Approximately 300μg of DNA in a volume of 130μl was sheared to a size of 300-

500bp using a Covaris S2 focused ultrasonicator (Intensity: 4; Fill level: 12; Duty cycle: 10%; Cycle/burst: 200; Time: 2 cycles of 40sec each) followed by a 0.5X-0.9X AMPure XP bead (Beckman Coulter, cat # A63881/ A63882) size selection. Biotin was removed from unligated ends with T4 DNA polymerase mix (1X NEBuffer 2, 0.05mM dATP, 0.05mM dGTP, 15U of T4 DNA polymerase (NEB, M0203)) at 20°C for 4hrs followed by streptavidin bead-based biotin pulldown (ThermoFisher Scientific, cat # 65001) in 1X streptavidin binding buffer (5mM Tris-HCl [pH 7.5], 0.5mM EDTA, 1M NaCl). Beads were washed 5 times in streptavidin tween washing buffer (5mM Tris-HCl [pH 7.5], 0.5mM EDTA, 1M NaCl, 0.05% Tween 20) and DNA was eluted off the beads twice with 50μl biotin elution buffer (10mM biotin solution [biotin solubilised in 0.1N NaOH], 20mM Tris-HCl [pH 8.0]) at room temperature for 30min followed by a 1X AMPure XP bead cleanup. KAPA Hyper Prep kit (Kapa Biosystems, cat # KK8502) was used for end-repair, A-tailing, adapter ligation with TruSeq index adapters, and 12 cycles of PCR amplification, all according to manufacturer's instructions. Previously described custom-designed biotinylated 120-mer RNA baits targeting Ensembl promoters of protein coding and non-coding transcripts were used to capture biotinylated and adapter-ligated promoter-associated Hi-C DNA from myometrium and leiomyoma samples using Agilent SureSelect XT kit (Agilent technologies, cat # G9611A) according to manufacturer's instructions [363]. TS universal blocker and TS index-specific blocker (IDT, xGen standard blocking oligos) were used in place of the SureSelect ILM Indexing Block 3. 6 cycles of post-capture, on-bead PCR amplification was performed and paired-end sequencing of all CHi-C libraries was performed on the Illumina NextSeq 500 platform.

2.7. General cell culture

All cells were cultured at 37°C in a humidified cell culture incubator containing 5% CO_2. Media was changed every 48 hours and cells were passaged at 90% confluence. HEK 293T/17 cells (ATCC, CRL-11268) were grown in high glucose-containing Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher, cat # 11965092) supplemented with 10% foetal bovine serum (FBS, ThermoFisher cat # 26149079). Human uterine smooth muscle cells (HUtSMC) were obtained from the American Type Culture Collection (ATCC, cat # PCS-460-011) and grown in vascular smooth muscle cell complete media (ATCC, cat #s PCS-100-030 and PCS-100-042). Myometrium smooth muscle cells were grown in smooth muscle cell culture complete media (SmGM-2 smooth muscle cell growth medium BulletKit [Lonza, cat # CC-3182], 1% Penicillin-Streptomycin).

2.8. RNAi-mediated silencing of AP-1 in myometrium primary cells

Short hairpin sequences targeting AP-1 subunits *JUN*, *JUNB*, *JUND*, *FOS,* and *FOSB* were obtained from the RNAi consortium (TRC). Short hairpin oligonucleotides (Integrated DNA Technologies) targeting AP-1 subunits and non-silencing control were cloned into a modified pLVX-IRES-mCherry vector (Clontech, cat # 631237) with bicistronic transcription driven by the phosphoglycerate kinase (PGK) promoter instead of the cytomegalovirus immediate early (CMV IE) promoter. EcoRI and BamHI restriction sites were used to integrate hairpin sequences into the vector and transformed into DH5α chemically competent cells overnight in a 37°C incubator using LB-ampicillin plates (100μg/ml ampicillin). Isolated colonies were grown for 16 hours in LB broth containing ampicillin (100μg/ml ampicillin), followed by miniprep according to manufacturer's instructions (Qiagen, cat # 27106). Lentivirus production was carried out in

HEK 293T/17 cells (ATCC, CRL-11268) with lipofectamine co-transfection of pMD2.G, psPAX2, and shRNA-containing lentiviral construct, a total of 18μg DNA at a ratio of 3:2:1, according to manufacturer's instructions (ThermoFisher, cat # 11668019). Virus was concentrated 10-fold using polyethylene glycol according to manufacturer's instructions (Takara Bio, cat # 631231). 1.0×10^6 myometrium primary cells collected from each patient ($n = 2$ biological replicates) and cultured in $10cm^2$ dishes with SmGM-2 muscle cell media, were each transduced with 100μl of either non-silencing control or AP-1 gene silencing concentrated virus and 6μg/ml of polybrene for 18 hours. This amount of virus was experimentally determined to be the minimum amount required to give > 80% transduction efficiency as determined by mCherry fluorescence. Fresh SmGM-2 muscle cell media was then added and cells were allowed to grow for 3-4 days. Cells were then harvested and RNA extracted using the RNeasy mini kit according to manufacturer's instructions (Qiagen, cat # 74104). Gene knockdown was confirmed by reverse transcription of purified RNA using qScript cDNA supermix (QuantaBio, cat # 95048-025) followed by quantitative PCR using SYBR Green PCR mix (ThermoFisher, cat # 4368708) using a Real-time PCR amplification system (ThermoFisher, cat # 4471134). Gene knockdown was also confirmed by western blot of JUN, JUNB (Bethyl, cat #s A302-958A, A302-704A respectively), JUND (Abcam, cat # ab28837), FOS and FOSB (ThermoFisher, cat #s MA5-15055, MA5-15056 respectively). Purified RNA was prepared for sequencing using the KAPA Stranded RNA-seq with RiboErase kit according to the manufacturer's instructions (Kapa Biosystems, cat # KK8483). Paired-end sequencing of all libraries was performed on the Illumina NextSeq 500 platform.

2.9. CRISPR/Cas9 mediated silencing of AP-1 in human uterine smooth muscle cells (HUtSMCs)

Custom CRISPR-Cas9 crRNAs targeting AP-1 subunits *JUN*, *JUNB*, *JUND*, *FOS*, and *FOSB* gene loci were obtained from Integrated DNA Technologies and resuspended in nuclease-free duplex buffer (IDT, cat # 11-01-03-01) to a concentration of 100μM. crRNAs were combined with 100μM tracrRNA (IDT, cat # 1072533), also resuspended in nuclease-free duplex buffer, in a 1:1 ratio to a final concentration of 44μM, incubated at 95°C for 5min and allowed to cool. SpCas9 enzyme (IDT, cat # 1081058) was diluted to 36μM with Neon resuspension buffer R (ThermoFisher, cat # MPK1025) and combined with the crRNA:tracrRNA AP-1 mixes in a 1:1 volume ratio and incubated at room temperature for 20min, after which electroporation enhancer (IDT, cat # 1075916) was added to a final concentration of 1.8μM.

HUtSM cells that were 90% confluent were trypsinized for 4min, quenched in FBS-containing media and spun down at 300 x *g* for 5min at room temperature. Centrifuged cells were resuspended in phosphate buffered saline and counted using an automated cell counter (ThermoFisher, cat # AMQAF1000) or a hemocytometer. Cells were then centrifuged at 300 x *g* for 5min at room temperature and resuspended at a concentration of 15,000 cells/μl in Neon resuspension buffer R. 135,000 cells were combined with crRNA:tracrRNA:cas9:electroporation enhancer mix targeting either a non-silencing control or AP-1 (*JUN*, *JUNB*, *JUND*, *FOS*, and *FOSB*) and electroporated in a 10μl Neon pipette tip (pulse voltage = 1,500V, pulse width = 10ms, # of pulses = 3) using a Neon electroporation system (ThermoFisher, cat # MPK5000) according to manufacturer's instructions. Electroporation of 135,000 cells in a 10μl Neon pipette

tip was repeated ten times and 1.35 x 10^6 electroporated cells were then cultured in smooth muscle cell complete media.

2.10. Verification of CRISPR/Cas9 mediated NHEJ in HUtSMCs

Cells were harvested and resuspended in 200μl of QuickExtract DNA solution (Epicentre, cat # QE09050) and incubated for 10min at 65°C followed by 95°C for 10min. 5μl of genomic DNA-containing QuickExtract solution was used to PCR amplify genomic regions around predicted AP-1 CRISPR cut sites. Gene editing was confirmed by sequence trace decomposition of sanger-sequenced PCR-amplified DNA [374].

2.11. RNA extraction followed by sequencing of AP-1 depleted HUtSMCs

Cells were harvested and RNA extracted using the RNeasy mini kit according to manufacturer's instructions (Qiagen, cat # 74104). Purified RNA was prepared for sequencing using the KAPA mRNA HyperPrep kit according to the manufacturer's instructions (Kapa Biosystems, cat # KK8580). Paired-end sequencing of all libraries was performed on the Illumina NextSeq 500 platform.

2.12. ChIP-sequencing of AP-1 depleted HUtSMCs

16% paraformaldehyde was added to full media-containing HUtSMC plates to a final concentration of 1% paraformaldehyde and incubated at room temperature for 10min. Cross-linked cells were then quenched with 0.125M glycine solution for a further 10min, followed by multiple washes in cold phosphate buffered saline containing 1X protease inhibitor cocktail (Roche, cat # 4693132001). Cells were then centrifuged for 5min at 1,000 x g and then resuspended in ChIP lysis buffer 1 (50mM HEPES-KOH [pH 7.6], 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% IGEPAL CA-630, 0.25% Triton X-100, 1X Protease inhibitor cocktail

[Roche, cat # 4693132001]) followed by end-over-end rotation at 4°C for 15min and then centrifuged for 5min at 1,000 x *g,* also at 4°C. Samples were resuspended in ChIP lysis buffer 2 (10mM Tris-HCl [pH 8.0], 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 1X Protease inhibitor cocktail [Roche, cat # 4693132001]) followed again by end-over-end rotation at 4°C for 15min and sample recovery by centrifugation at 1,000 x *g* at 4°C for 5min. Samples were then resuspended in ChIP lysis buffer 3 (10mM Tris-HCl [pH 7.5], 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% sarkosyl) and sonicated in an ice water bath (Misonix, setting 6 [~6 W power output], 12 cycles of 15sec on and 45sec off). Triton X-100 was added to a final concentration of 1% and samples were centrifuged at 20,000 x *g* at 4°C for 20min, with recovery of chromatin-containing supernatant. Solubilised chromatin concentration was measured by BCA assay according to manufacturer's instructions (ThermoFisher, cat # 23225). Approximately 500μg of chromatin and 4μg of antibody against H3K27Ac (Active motif, cat # 39685) was used for overnight immunoprecipation of histones at 4°C with end-over-end rotation. 10μl of protein G magnetic beads (ThermoFisher, cat # 10004D) per μg of antibody was added and samples were incubated at 4°C for an additional 3 hours. Immunoprecipitated samples bound to beads were recovered using a magnetic rack (ThermoFisher, cat # 12321D), followed by 5 washes with ChIP-RIPA wash buffer (50mM HEPES-KOH [pH 7.6], 500mM LiCl, 1mM EDTA, 1% IGEPAL CA-630, 0.7% sodium deoxycholate) and once with NaCl containing TE buffer (10mM Tris-HCl [pH 8.0], 1mM EDTA, 50 mM NaCl), with magnetic recovery of immunoprecipitated samples bound to beads after each wash. DNA was eluted off the beads twice with 50μl ChIP elution buffer (0.1M NaHCO$_3$, 1% SDS) at 65°C for 20min with shaking. Cross-links were reversed with 300mM NaCl for 12-16 hours at 65°C followed by 20μg

of RNase A (Worthington, cat # LS002132) treatment at 37°C for 1 hour and then 2 hours at

55°C with 80μg of proteinase k (ThermoFisher, cat # 25530015) supplemented with 16.5mM

EDTA and 66mM Tris-HCl (pH 8.0). Reverse cross-linked DNA was isolated by phenol-

chloroform extraction with phenol/chloroform/isoamyl alcohol [25:24:1] (Sigma cat, # 77617)

using 5Prime phase lock gel (Quantabio, cat # 2302830) according to manufacturer's

instructions, followed by ethanol precipitation (2X volume of 100% ethanol, 0.1X volume of 3M

sodium acetate, and 2μl of glycogen) at -80°C for 1 hour. DNA was then pelleted by

centrifugation at 20,000 x g at 4°C for 20min, followed by one wash with 70% ethanol and

further centrifugation at 20,000 x g at 4°C for 10min. DNA pellet was dried and resolubilised in

50μl of elution buffer (10mM Tris-HCl, pH 8.0 - 8.5). KAPA Hyper Prep kit (Kapa Biosystems,

cat # KK8502) was used for end-repair, A-tailing, and adapter ligation with TruSeq index

adapters, all according to manufacturer's instructions. A 0.6X-0.8X AMPure XP bead (Beckman

Coulter, cat # A63881) size selection was carried out after adapter ligation to enrich for mono-

nucleosomal DNA fragments from the H3K27Ac ChIP followed by 12 cycles of PCR

amplification. Single-end sequencing of all ChIP libraries was performed on the Illumina

NextSeq 500 platform.

2.13. Immunoblotting

Cell culture plates containing cultured primary cells were washed in phosphate buffered saline.

Cells were then harvested in phosphate buffered saline containing protease inhibitor cocktail

(Roche, cat # 4693132001) and centrifuged for 5min at 1,000 x g. Cell pellet was resuspended in

200μl RIPA buffer (10mM Tris-HCl [pH 8.0], 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1%

sodium deoxycholate, 0.1% SDS, 1% triton X-100, 1X Protease inhibitor cocktail [Roche, cat #

4693132001]) and incubated on ice for 30min with intermittent shaking using a vortex mixer. This was followed by brief sonication in an ice water bath (Misonix, setting 6 [~6 W power output], 2 cycles of 15sec on and 45sec off) and then centrifugation at 20,000 x g at 4°C for 20min, with recovery of the whole cell lysate-containing supernatant. The concentration of the whole cell lysate was measured by BCA assay according to manufacturer's instructions (ThermoFisher, cat # 23225). Lysate was then mixed with β-ME containing 2X laemmli buffer in a 1:1 ratio and incubated at 98°C for 8min. 30μg of lysate was then loaded into a well of an 8-16% tris-glycine gel (ThermoFisher, cat # XP08160) and the gel was run in tris-glycine buffer containing 0.1% SDS. The protein in the gel was then transferred onto a nitrocellulose membrane (GE Healthcare, cat # 10600001) for 2 hours using a gel transfer apparatus in gel transfer buffer (0.5X tris-glycine, 20% methanol, 0.01% SDS) at 30V. After transfer, membrane was stained with Ponceau stain to ensure equal loading of protein and then de-stained with phosphate buffered saline containing 0.05% Tween (PBST). Membrane was then blocked for 30min with rocking using blocking milk solution (PBST containing 5% skim milk). Membrane was then incubated overnight with rocking in blocking milk solution containing primary antibody at 4°C. This was followed by one 10min membrane wash in reverse osmosis water and three 10min washes in PBST. The membrane was then incubated at room temperature for 1 hour with rocking in blocking milk solution containing the appropriate secondary antibody. This was followed by one 10min membrane wash in reverse osmosis water and three 10min washes in PBST. The membrane was then covered with chemiluminescent detection reagent (GE Healthcare, cat # RPN2232) for 2min and developed using chemiluminescence film (GE Healthcare, cat # 28906839) according to manufacturer's instructions.

2.14. General analysis and data processing

All ChIP, CHi-C, and RNA sequencing reads in FASTQ format were aligned to the GRCh38 human genome assembly obtained from Ensembl (release 90) and all transcript annotations were also obtained from Ensembl. The R package Gviz was used to visualise browser tracks of ChIP-seq, and RNA-seq BigWig files as well as myometrium and leiomyoma CHi-C q scores [375].

2.15. RNA-seq data processing and analysis

RNA sequencing reads from tissue samples (n = 15 patients) and cells (n = 3 HUtSMCs, n = 2 myometrium primary cells obtained from patients undergoing hysterectomies at Prentice women's hospital) were aligned to the GRCh38 human genome assembly using the STAR aligner (v2.5.3a) with default settings (--alignIntronMin 20 --alignIntronMax 500000) [376]. Uniform gene body coverage of reads was verified with the RSeQC python package (v2.6.4) [377]. RSeQC was also used to normalise BAM files obtained from STAR alignments to 1×10^7 reads when converting to wiggle format, which were then converted to BigWig files using the UCSC browser utility wigToBigWig. Counting of reads per gene was performed using featureCounts from the Subread package (v1.5.0; introns + exons count settings: -g gene_id -t gene -p -s 2, exon count settings: -g gene_id -t exon -p -s 2) and differential expression analysis of gene counts between myometrium and leiomyoma was performed with DESeq2 (v1.18.1), with the Wald test used to test for significance (p-value). The false discovery rate (adjusted p-value) was controlled for at < 0.01 [378,379]. RNA tissue clustering and differential gene expression heat maps were generated in R using pheatmaps (cran.r-project.org). All reads (introns + exons) mapping to an annotated feature were used to compare transcriptome profiles between samples whilst only exonic reads were used for differential gene expression analysis. TPMcalculator was used to

calculate transcripts per million (TPM) normalisation of tissue RNA-seq reads [380]. DESeq2

determined normalisation of growth factor counts per gene were qualitatively similar to TPM

normalised counts. Enriched gene ontology terms were identified using Metascape and scatter

plots of ontology terms were made using REVIGO [381,382]. Exon-intron split analysis (EISA) and

analysis of differential exon usage were performed as previously described [383,384].

2.16. ChIP-seq data processing and analysis

ChIP sequencing reads of tissue samples (n = 5 patients) were first adapter trimmed using

cutadapt (v1.15). Read quality of tissue samples and cells (n = 3) were then verified with FastQC

(www.bioinformatics.babraham.ac.uk) [385]. Reads were then aligned to the GRCh38 human

genome assembly using bowtie (v1.2.2; settings: --best -m 1) [386]. H3K27Ac peak calling and

motif enrichment were performed with HOMER (v4.10.3, homer.ucsd.edu; -tbp 1 -F 4 -style

histone). Transcription factor peak was performed with the following parameters: -tbp 1 -F 3 -

style factor. Differential enrichment of peaks was determined using the R package DiffBind [387]

(www.bioconductor.org). First, DiffBind was used to derive a consensus myometrium and

leiomyoma peak set for each histone or factor ChIP from the individual HOMER-called peak

sets of each biological replicate. Differential peak analysis on the consensus peak set was then

performed with a peak being identified as differentially enriched or depleted at an FDR cutoff <

0.05 and a fold change > |2|. Promoters are defined as regions from -1000bp to +100bp of the

transcriptional start site and enhancers regions are H3K27Ac peaks that fall outside promoter

regions and are further subdivided into H3K27Ac peaks found in exonic, intronic, and intergenic

regions. Pearson correlation heat maps of biological replicates were also drawn using DiffBind.

Reads per genomic content normalisation (RPGC, 1x normalisation) of BAM files obtained from

Bowtie alignments was performed with DeepTools to produce BigWig files [388]. Heat maps and profile plots of ChIP-seq peaks were also plotted using DeepTools. Motif enrichment analysis at differential H3K27Ac regions was carried out using the HOMER function findMotifsGenome using default parameters (-size given -len 8,10,12,16,20).

2.17. Capture Hi-C data processing and analysis

Promoter capture Hi-C sequencing reads ($n = 5$ patients) were processed using the Hi-C User Pipeline (HiCUP, v0.5.10), with the bowtie aligner (v1.2.2) being used to align the reads to the GRCh38 human genome assembly [386,389]. The capture Hi-C analysis of genomic organisation package (CHiCAGO) was used to score CHi-C contact strength, with HiCUP processed BAM files being used as input files for the CHiCAGO pipeline [390]. Enhancer-promoter (E-P) and promoter-promoter (P-P) interactions in myometrium and leiomyoma samples with a CHiCAGO score > 5 were considered to be valid contacts. The edgR package (v3.20.9) was used to identify significantly altered contacts at an FDR < 0.1 [391]. Only contacts with a mean read count > 10 for each tissue type were used in the edgeR analysis. Differential contacts identified by edgR analysis were validated with Chicdiff, a computational pipeline designed specifically for detecting differential chromosomal interactions in Capture Hi-C data [392]. Heat maps of differential contacts were generated in R using pheatmaps (cran.r-project.org). *Hin*dIII digest of DNA in CHi-C sample preparation was leveraged as a means to assign overlap between enhancer-promoter contacts and H3K27 acetylated regions. H3K27Ac peaks that overlapped the same *Hin*dIII fragment as the distal other end of an enhancer-promoter contact were assigned to the same *Hin*dIII fragment/region.

2.18. Data availability

All high-throughput sequencing data that support the findings of this study have been made available through the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) data repository and can be accessed via accession GSE128242.

CHAPTER 3. RESULTS

3.1. Transcriptome profiling of fibroid RNA reveals changes in transcriptional regulation of genes associated with key biological processes in uterine leiomyomas

RNA isolation followed by massively parallel sequencing (RNA-seq) was used to examine the transcriptome profiles of normal myometrium (WT) and matched leiomyoma (G44D/S) tissue obtained from 15 women (Fig. 3.1.1). A high degree of similarity between biological replicates of myometrium transcriptome profiles was seen, with a similar observation among biological replicates of leiomyoma tissue samples. Hierarchical clustering of all RNA-seq datasets highlights clustering primarily by disease state (Fig. 3.1.2a). Significantly, principal component analysis of the most variable genes revealed that 43% of the variance (PC1) is explained by the disease state, with biological replicates co-segregating based on tissue type (Fig. 3.1.2b). This suggests that the changes in gene expression between normal and *MED12* mutant disease tissue types are primarily attributable to biological pathways that are important for the development and maintenance of the leiomyoma disease state.

Differential gene expression analysis identified 5,831 dysregulated genes at a false discovery rate of 0.01 (Fig. 3.1.3). Protein-coding genes constituted an overwhelming majority of the dysregulated transcriptional output, accounting for approximately 79% ($n = 4,597/5,831$) of all differentially expressed genes (Fig. 3.1.4). Long non-coding RNAs, antisense RNAs, and pseudogenes together accounted for approximately 19.7% of differentially expressed transcripts. However, the number of small non-coding RNAs that are dysregulated in leiomyomas is likely underrepresented, as the extraction and library preparation protocol does not recover these RNAs. 2,965 genes exhibited a > 2-fold change in expression between myometrium and leiomyoma samples, whilst 1,014 genes were differentially expressed at > 4-fold (Fig. 3.1.5).
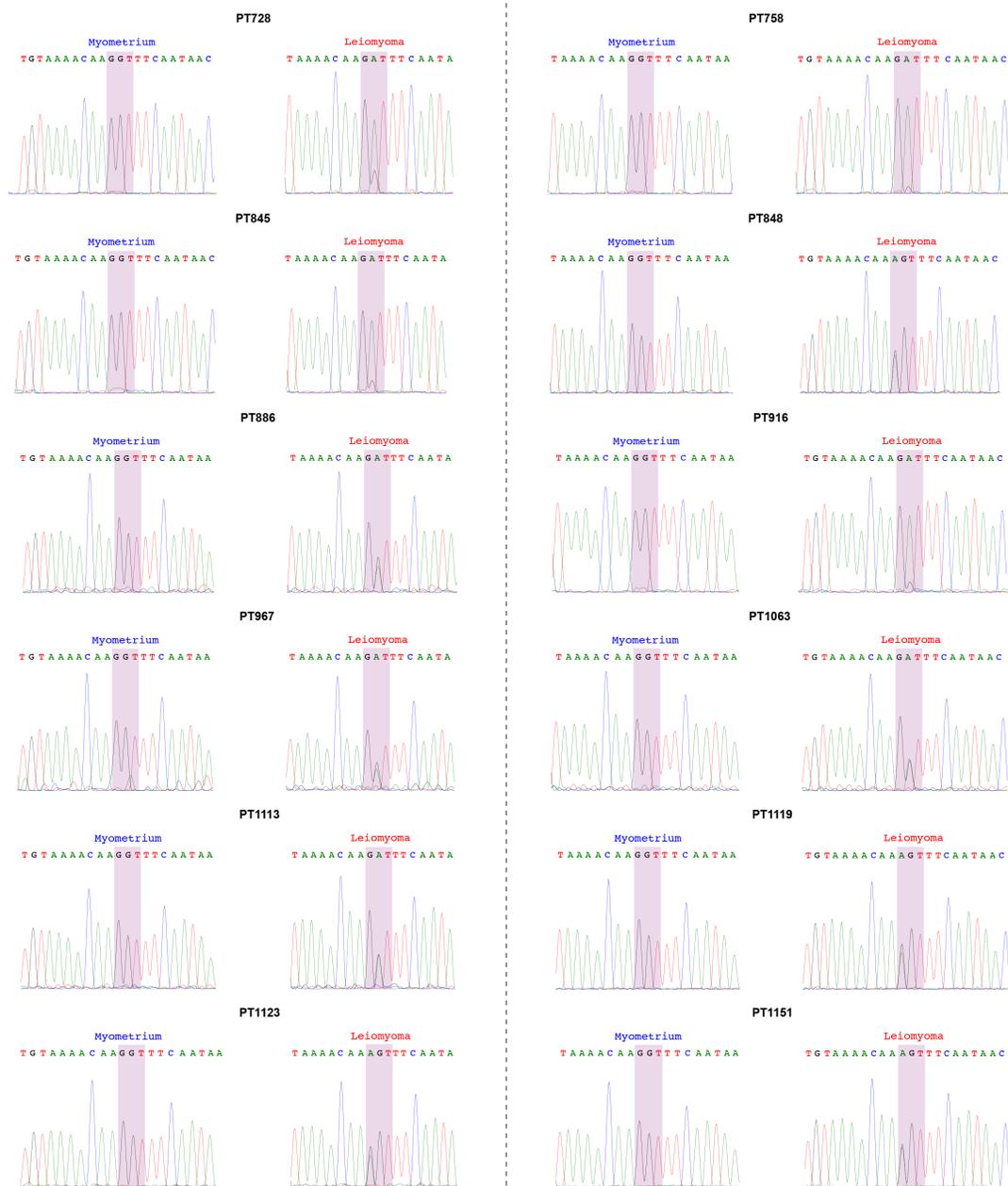
Figure 3.1.1. Identification of *MED12* mutant patient tissue samples. Sanger sequencing of gDNA or cDNA extracted from myometrium and leiomyoma patient tissue samples. Myometrium samples were confirmed to be wild type for *MED12* mutations and leiomyomas verified to have G44D or G44S mutations in *MED12*.

88
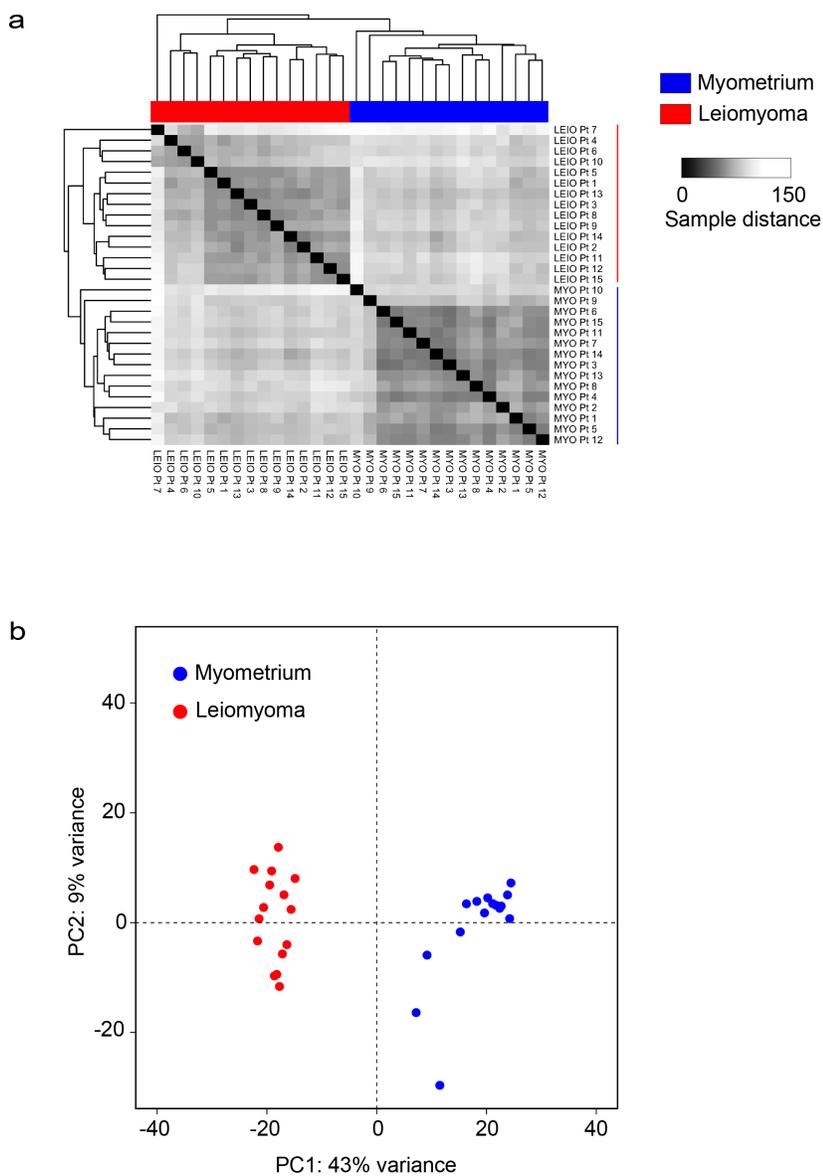
Figure 3.1.2. Patient tissue transcriptome profiles. (**a**) Heat map representing tissue sample Euclidean distance matrix clustering of RNA sequencing profiles for 15 patients. (**b**) Principal component analysis plot of the top 500 most variable genes in myometrium and leiomyoma tissue samples. First two principal components are plotted.
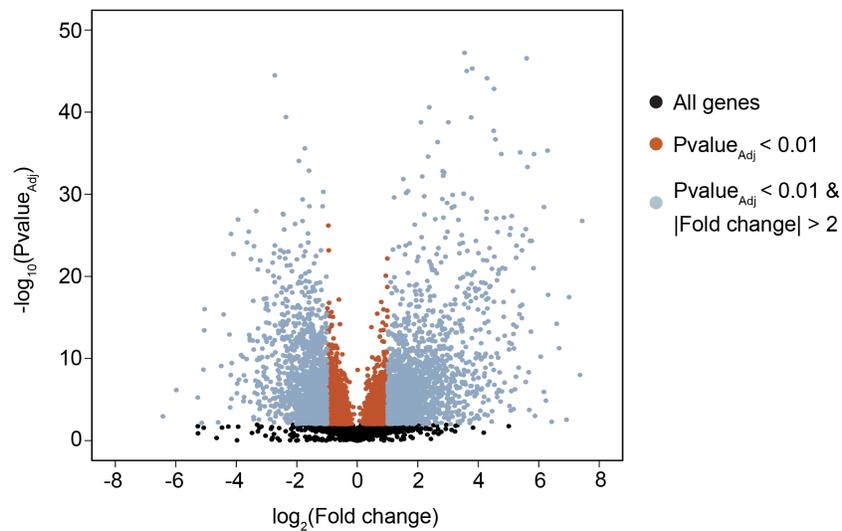
Figure 3.1.3. Volcano plot of differentially expressed genes between myometrium and leiomyoma. Differentially expressed genes (sienna, $n = 5,831$) and differentially expressed genes with > 2-fold change (grey, $n = 2,965$) are highlighted (FDR < 0.01, BH-corrected Wald test). Unchanged transcripts (FDR > 0.01) are shown in black. *P* values are truncated at 1×10^{-50} for visualisation purposes.

protein coding (78.8%)   misc RNA (0.4%)
lncRNA (9.5%)   snRNA (0.3%)
antisense RNA (5.7%)   snoRNA (0.3%)
pseudogene (4.5%)   scaRNA (0.1%)
miRNA (0.3%)   rRNA (0.1%)

Total = 5831
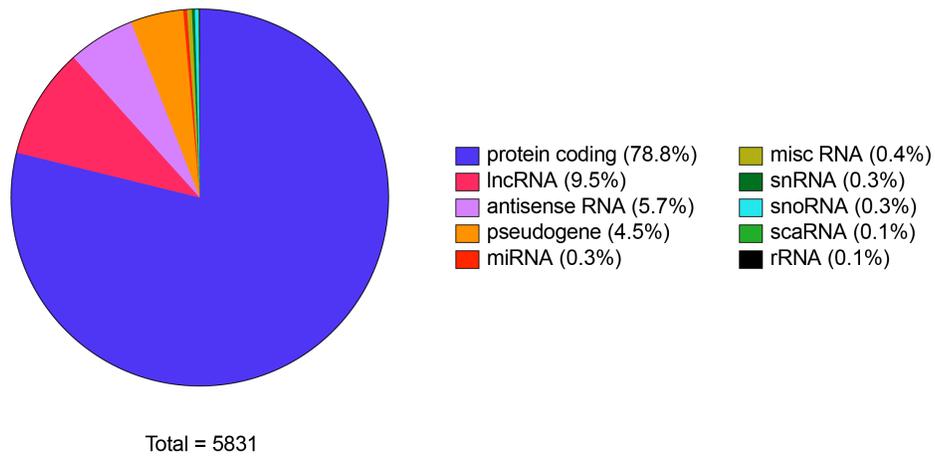
Figure 3.1.4. Protein-coding genes account for the majority of dysregulated genes in uterine leiomyomas. Pie chart of transcript biotypes for differentially expressed genes in uterine leiomyomas (n = 5,831).

Figure 3.1.5. Heat map of differentially expressed genes with > 2-fold change in uterine tissue samples. Gene expression levels in myometrium (blue) vs. leiomyoma (red) relative to the mean expression are shown as row z scores (*n = 2,965 genes*).
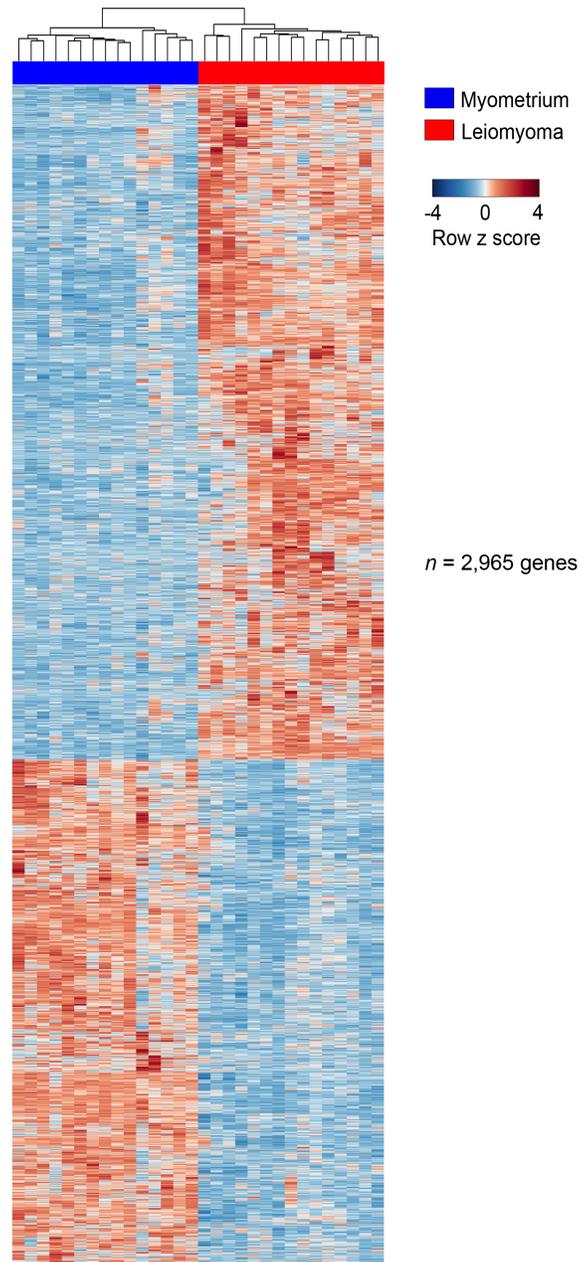
A significant number of dysregulated genes in leiomyoma samples are associated with extracellular matrix formation, organisation and degradation, which are key up-regulated biological processes in uterine fibroid disease pathogenesis (Fig. 3.1.6 and Fig. 3.1.7) [1]. In total, 325 matrisome genes were dysregulated, with 216 of these constituting core ECM genes (Fig. 3.1.8a, b and Fig. 3.1.9a, b, Supplementary table 7). Collagen genes were highly dysregulated, with an increased expression observed in the majority of dysregulated collagens (Fig. 3.1.10). In addition to collagens, a significant number of fibrous proteins and glycoproteins were also dysregulated in leiomyomas (Fig. 3.1.11). Notable inclusions were fibronectin, laminins (LAMA3, LAMC3, LAMB1, -2, and -4), EMILINs (EMILIN2 and -3), fibulins (FBLN1, -2, and -3), and fibrillins (FBN2 and -3). Proteoglycans were also dysregulated, with the hyalectan group of glycoproteins well represented in this group (Fig. 3.1.12).

Genes coding for matrix-associated proteins involved in ECM degradation were also highlighted in the list of dysregulated genes in leiomyoma tissue samples (Fig. 3.1.13). Proteolytic metalloproteinases belonging to MMP, ADAM and ADAMTS families were well represented in this group, with a noticeable absence of TIMPs. Counter intuitively, a significant number of metalloproteinases demonstrated an increased gene expression in leiomyoma tissue samples relative to myometrium, suggesting that the increase in ECM formation and deposition observed in fibroids is not due to a down regulation in matrix degrading enzymes. Also surprising, lysyl oxidases and lysyl hydroxylases, which regulate matrix stiffness, were in general not perturbed in leiomyomas, suggesting that increased matrix rigidity observed in fibroids is regulated by an alternative mechanism.
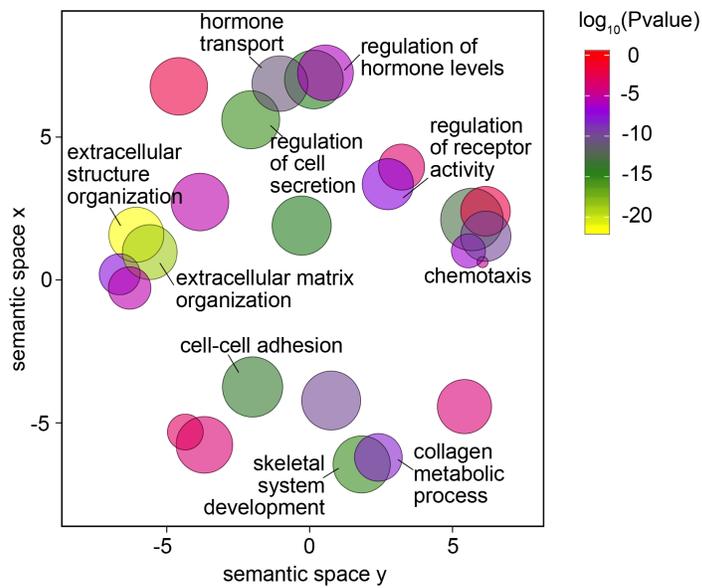
Figure 3.1.6. Differentially expressed genes in myometrium vs. leiomyoma are involved in extracellular matrix formation, organisation and degradation. Scatter plot of confidence scores for enriched gene ontologies associated with differentially expressed genes, with ontologies clustered by functional similarity in the semantic space.
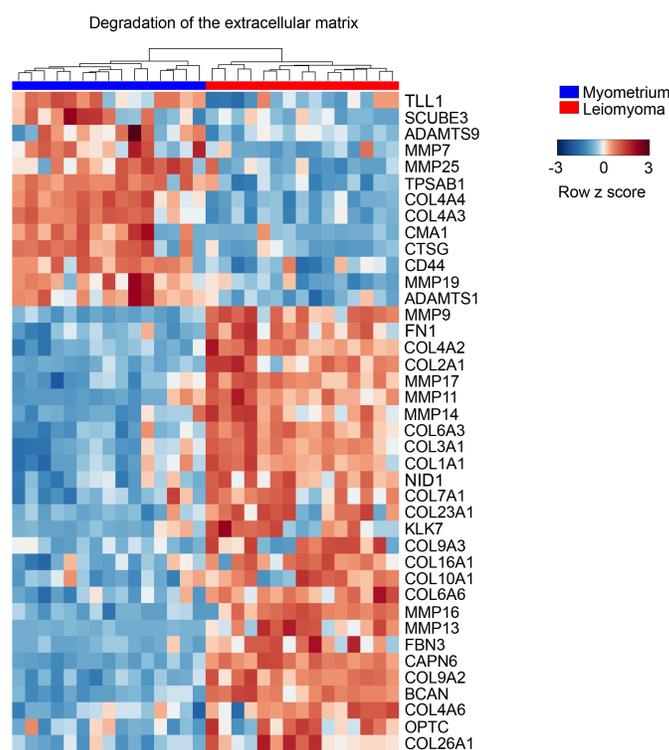
Figure 3.1.7. Hierarchically clustered heat map of differentially expressed genes associated with ECM degradation. Dysregulated genes associated with the degradation of the extracellular matrix (R-HSA-1474228) gene ontology term are shown (Supplementary Table 7). Gene expression levels in myometrium (blue) vs. leiomyoma (red) relative to the mean expression are shown as row z scores.
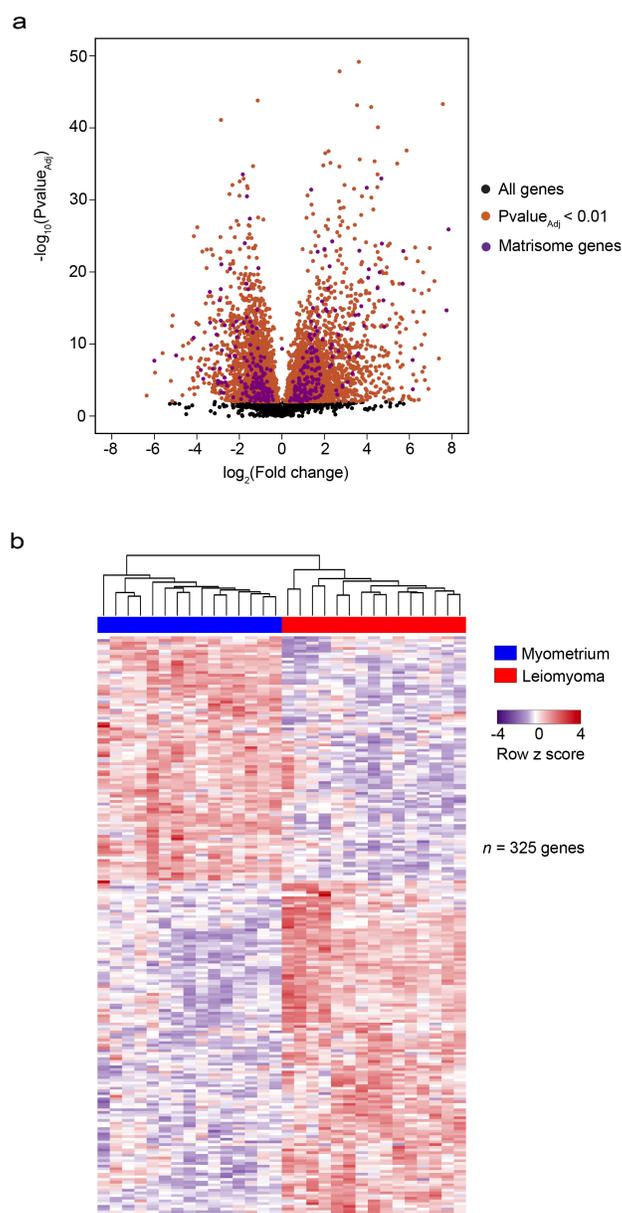
Figure 3.1.8. Dysregulated matrisome associated genes in uterine leiomyomas. (a) Volcano plot of transcript expression levels in patient myometrium and leiomyoma samples. All differentially expressed genes (sienna, $n = 5{,}831$) and differentially expressed genes associated with the core matrisome (GO:M5884) ontology term (purple, $n = 325$), which comprises extracellular matrix and extracellular matrix associated genes, are highlighted (FDR < 0.01, BH-corrected Wald test). Unchanged transcripts (FDR > 0.01) are shown in black. *P* values are truncated at 1×10^{-50} for visualisation purposes. **(b)** Hierarchically clustered heat map representing the volcano plot in (a) of all differentially expressed core matrisome genes. Gene expression levels relative to the mean expression are shown as row z scores ($n = 325$ genes).
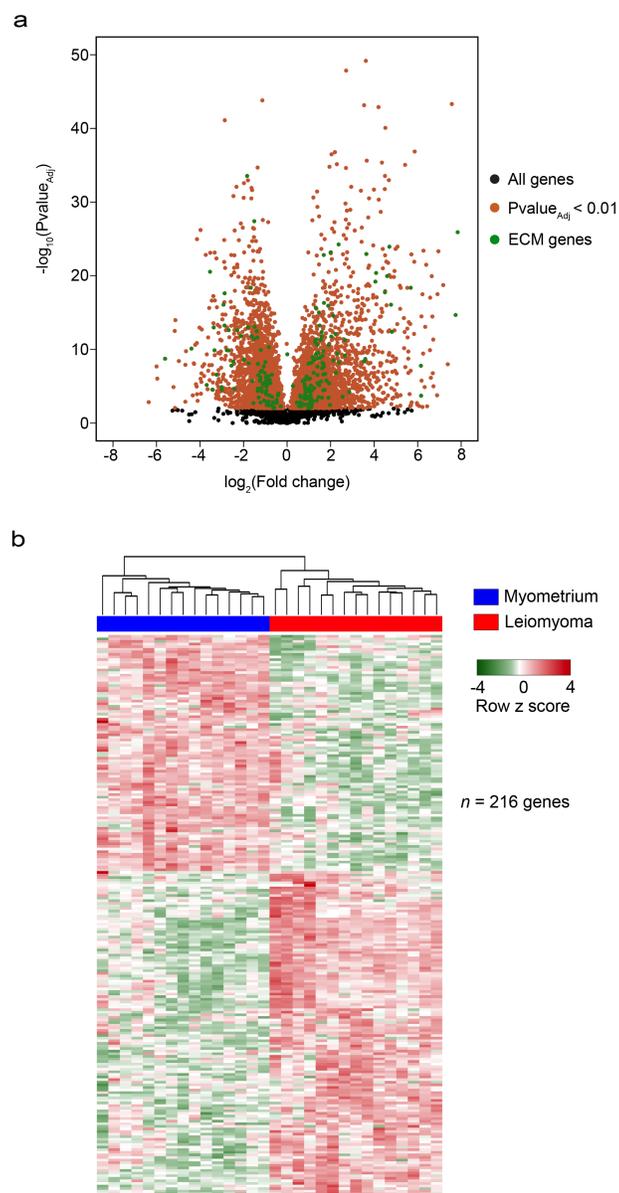
Figure 3.1.9. Dysregulated core extracellular matrix genes in uterine leiomyomas. (a) Volcano plot of expression levels in patient myometrium and leiomyoma samples. All differentially expressed genes (sienna, $n = 5{,}831$) and differentially expressed genes associated with the production and regulation of the extracellular matrix (green, $n = 216$), which is a subset of the matrisome ontology term, are highlighted. (FDR < 0.01, BH-corrected Wald test). Unchanged transcripts (FDR > 0.01) are shown in black. *P* values are truncated at 1×10^{-50} for visualisation purposes. **(b)** Hierarchically clustered heat map representing the volcano plot in (a) of all differentially expressed genes associated with the production and regulation of the extracellular matrix (ECM). Gene expression levels relative to the mean expression are shown as row z scores ($n = 216$ genes).

Figure 3.1.10. Hierarchically clustered heat map of differentially expressed collagen genes in uterine leiomyomas. Gene expression levels relative to the mean expression are shown as row z scores.

Figure 3.1.11. Hierarchically clustered heat map of differentially expressed genes coding for glycoproteins and fibrous proteins. Gene expression levels relative to the mean expression are shown as row z scores.

Figure 3.1.12. Hierarchically clustered heat map of differentially expressed genes coding for proteoglycans. Gene expression levels relative to the mean expression are shown as row z scores.

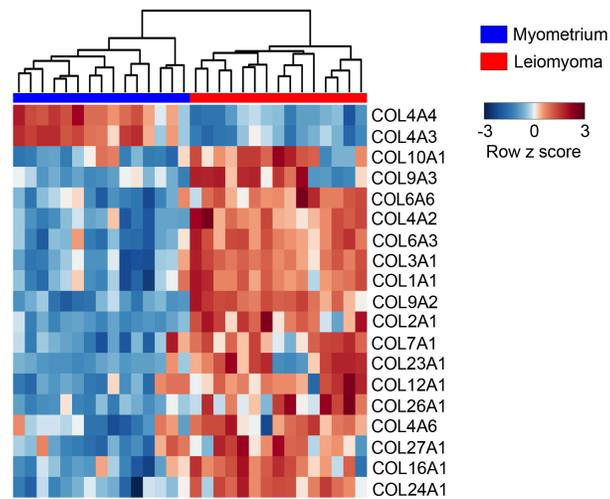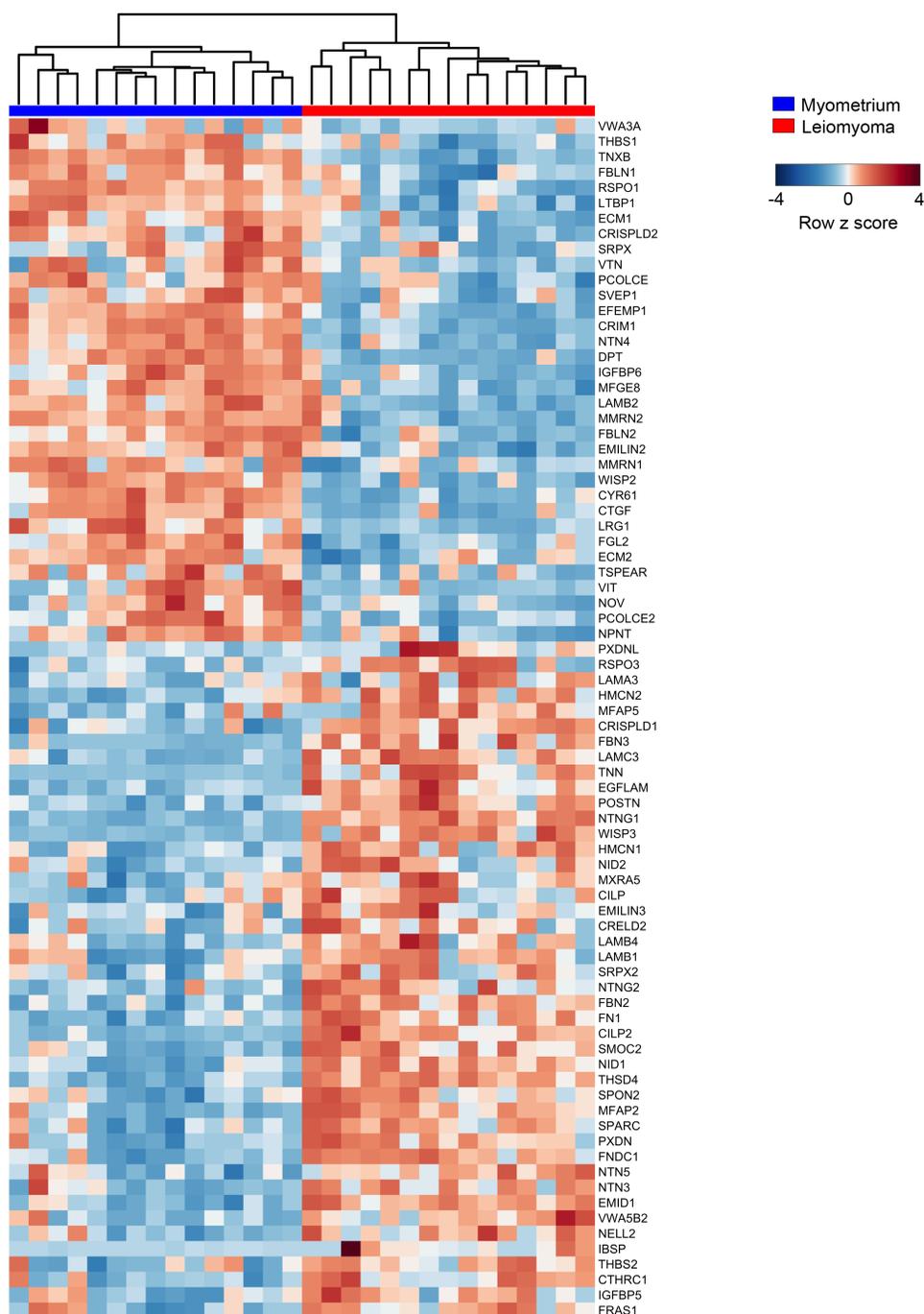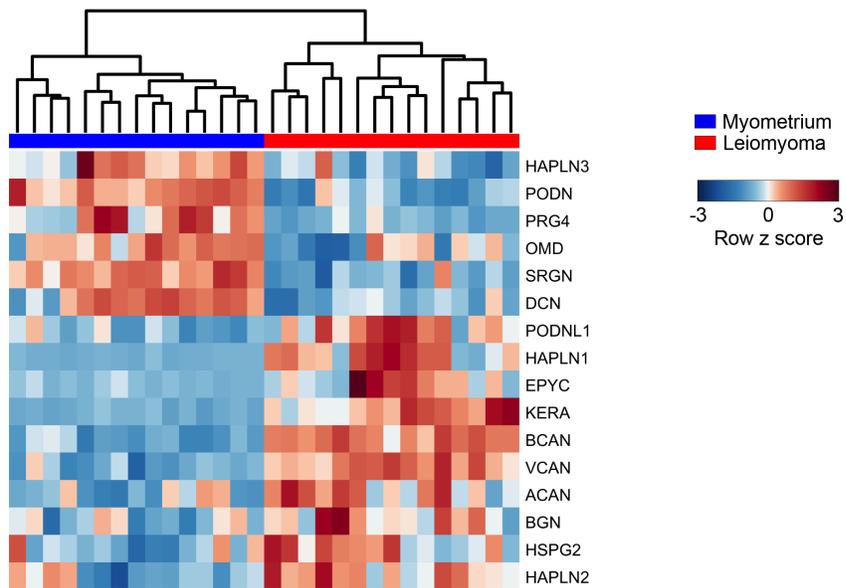**Figure 3.1.13. Hierarchically clustered heat map of differentially expressed genes coding for regulators of core ECM genes.** Gene expression levels relative to the mean expression are shown as row z scores.

Given previous studies outlining the importance of growth factors in leiomyoma pathogenesis, growth factor gene expression in leiomyomas compared to normal myometrium was investigated (Fig. 3.1.14). In the epidermal growth factor family, ErbB3 receptor expression was down regulated in leiomyomas, although it was expressed at low levels in both myometrium and leiomyomas tissues. ErbB3 receptor down regulation may have been in response to an increase in expression of neuregulins [393]. In addition to ErbB3, ErbB1 also saw a modest decrease in expression whilst betacellulin and HB-EGF were other members of the epidermal growth family of factors found to be dysregulated in leiomyomas (Fig. 3.1.14). However, previously reported changes in EGF expression were not observed [133]. In concordance with previously reported data, Insulin growth factor II (IGF-II) was upregulated in leiomyomas samples [146]. Although IGFBP5 was differentially expressed, changes were modest whilst neither IGF receptors II nor I were found to be dysregulated.

Twelve fibroblast growth family members are expressed in myometrium and leiomyoma samples, namely FGF1, -2, -7, FGF9 – FGF14, and FGFs -17, -18, and -22 (Fig. 3.1.14 and Fig. 3.1.15a). FGF2 and FGF7 are highly expressed, whilst FGF1, -11 and -13 are moderately expressed, with the rest being expressed at low levels. FGF2, -12 and -13 were down regulated in leiomyomas whilst FGF1 was increased in expression. In addition, FGF receptor 3 saw a modest change in expression (Fig. 3.1.15b). These results were generally consistent with previously reported data [132,139,140]. PDGF/VEGF family members PDGFB, -C and -D as well as VEGFA were also dysregulated in leiomyomas although no significant change was seen in receptor expression (Fig. 3.1.14 and Fig. 3.1.16a). Neither PDGF receptors A or B were perturbed, however all three VEGF receptors were down regulated in leiomyomas (Fig. 3.1.16b).
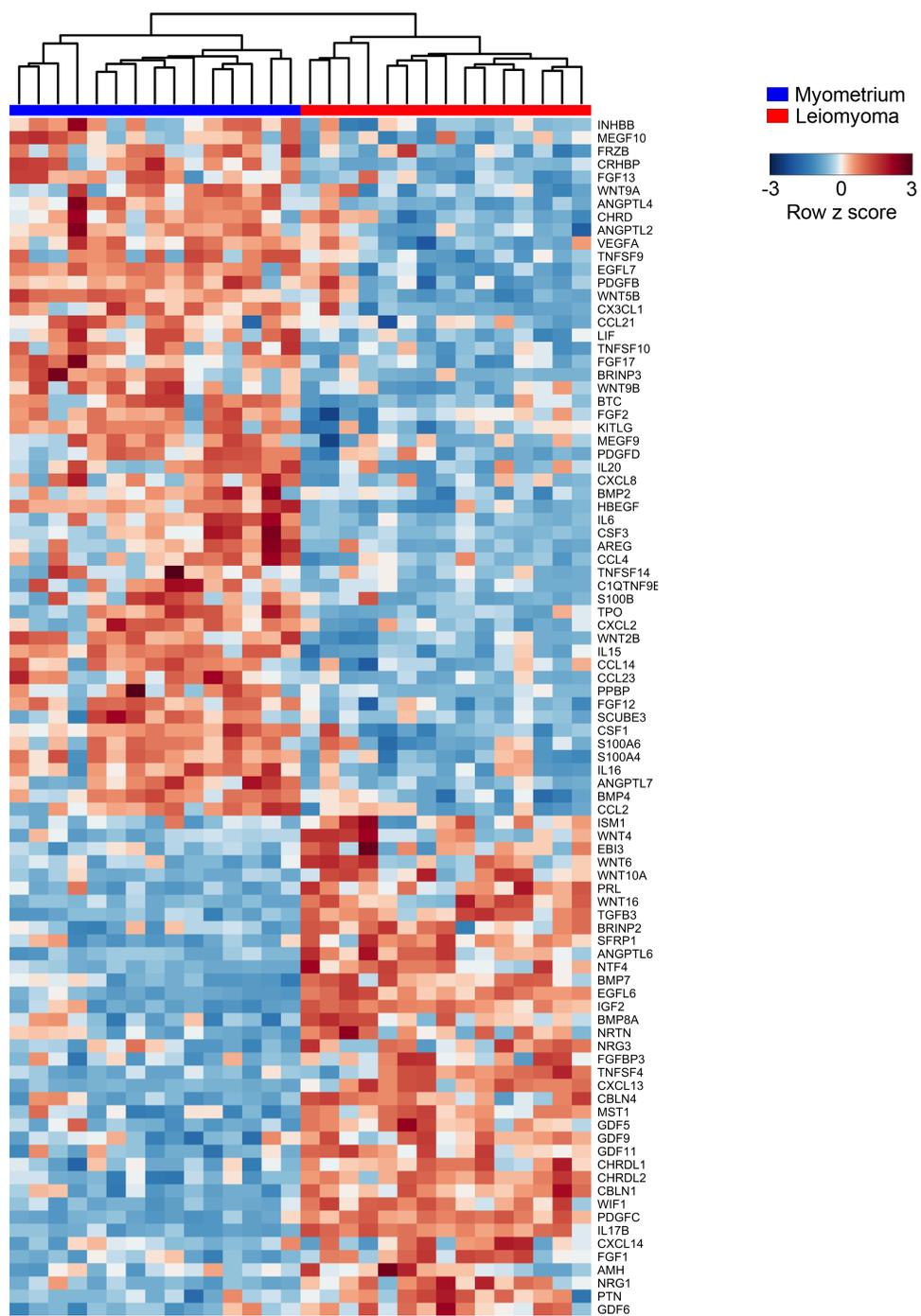
Figure 3.1.14. Hierarchically clustered heat map of differentially expressed genes coding for growth factors and other secreted factors. Gene expression levels relative to the mean expression are shown as row z scores.
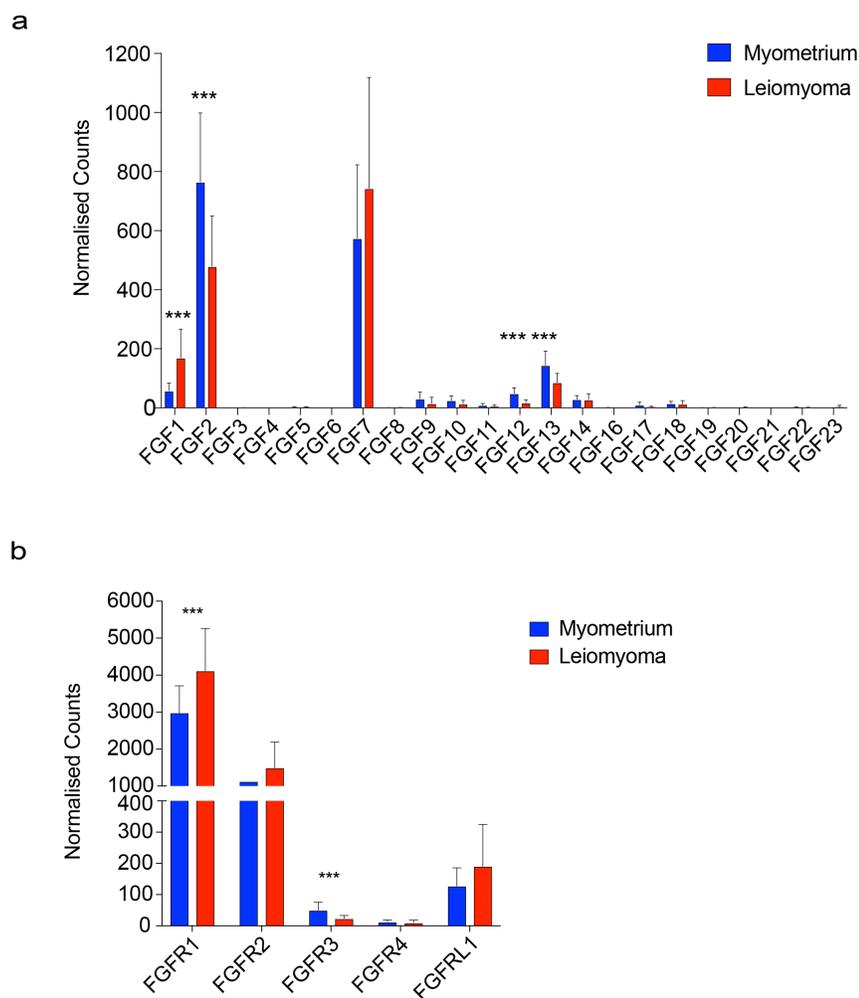
Figure 3.1.15. Fibroblast growth factor family gene expression in myometrium and leiomyoma tissue. (**a**) Bar graph of fibroblast growth factor family member gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples ($n = 15$ patients). (**b**) Bar graph of fibroblast growth factor family receptor gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples ($n = 15$ patients). *** Indicates genes found to be differentially expressed at an FDR < 0.01.

Figure 3.1.16. Platelet derived growth factor/ Vascular endothelial growth factor family gene expression in myometrium and leiomyoma tissue. (**a**) Bar graph of PDGF/VEGF family member gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (n = 15 patients). **(b)** Bar graph of PDGF/VEGF family receptor gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (n = 15 patients). *** Indicates genes found to be differentially expressed at an FDR < 0.01.

Consistent with previous reports, TGF-β3 was highly upregulated in uterine leiomyomas, although TGF-β receptor II was down regulated, which may blunt the effect of increased TGF-β3 mediated signalling (Fig. 3.1.14, Fig. 3.1.17a, Fig. 3.1.18a). Interestingly, seven growth factors belonging to the BMP and GDF subfamilies were dysregulated in leiomyoma samples (Fig. 3.1.14, Fig. 3.1.17b). Altered expression of multiple type I and type II receptors that are bound by BMPs and GDFs was also observed (Fig. 3.1.18a). In addition, SMAD1 and SMAD8/9, R-SMADs activated by BMP and GDF signalling, are also dysregulated in uterine leiomyomas (Fig. 3.1.18b). This may suggest that BMP and GDF subfamily signalling plays a more prominent role than TGF-β subfamily signalling in uterine leiomyoma disease pathogenesis.

Regulation of RNA levels occurs primarily at transcription, which includes epigenetic control of the RNAPII-mediated transcription machinery. RNA levels are also regulated post-transcriptionally, through the regulation of RNA processing and degradation [394]. It has been demonstrated that differences in intronic read counts from RNA-sequencing experiments can serve as a measure of changes in transcriptional activity and even predict post transcriptionally regulated genes [383,395]. To investigate possible post transcriptionally driven processes in uterine leiomyomas, exon-intron split analysis (EISA) was carried out. Exonic and intronic changes between myometrium and leiomyoma tissue samples were highly correlated, indicating that gene expression changes seen in the disease state were primarily transcription-driven and not posttranscriptional (Fig. 3.1.19, $R = 0.89$). However, 331 genes were identified as possible targets of posttranscriptional regulation. Scanning of the 3' UTRs of genes identified seed sequences for mir-216 and mir-124a, which have both been demonstrated to play a role in
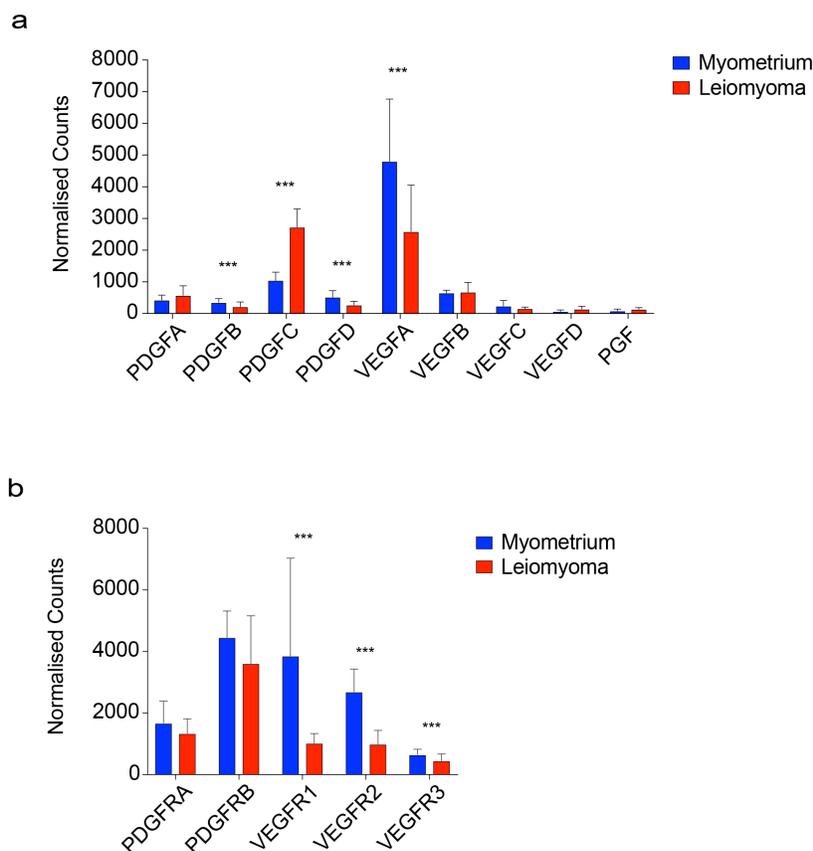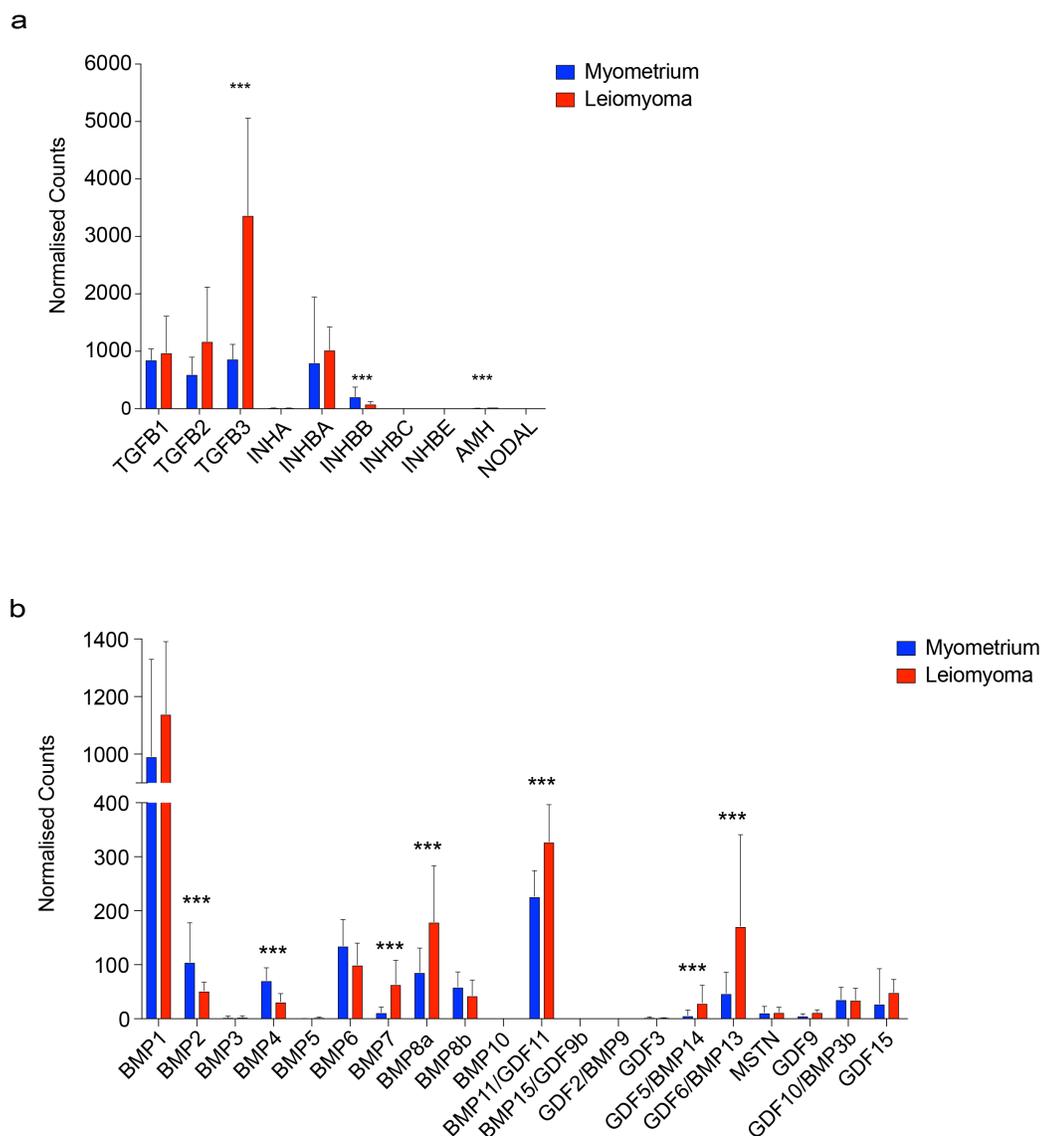
Figure 3.1.17. Transforming growth factor-β family gene expression in myometrium and leiomyoma tissue. (a) Bar graph of TGF-β subfamily member gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (*n* = 15 patients). **(b)** Bar graph of BMP/GDF subfamily member gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (*n* = 15 patients). *** Indicates genes found to be differentially expressed at an FDR < 0.01.
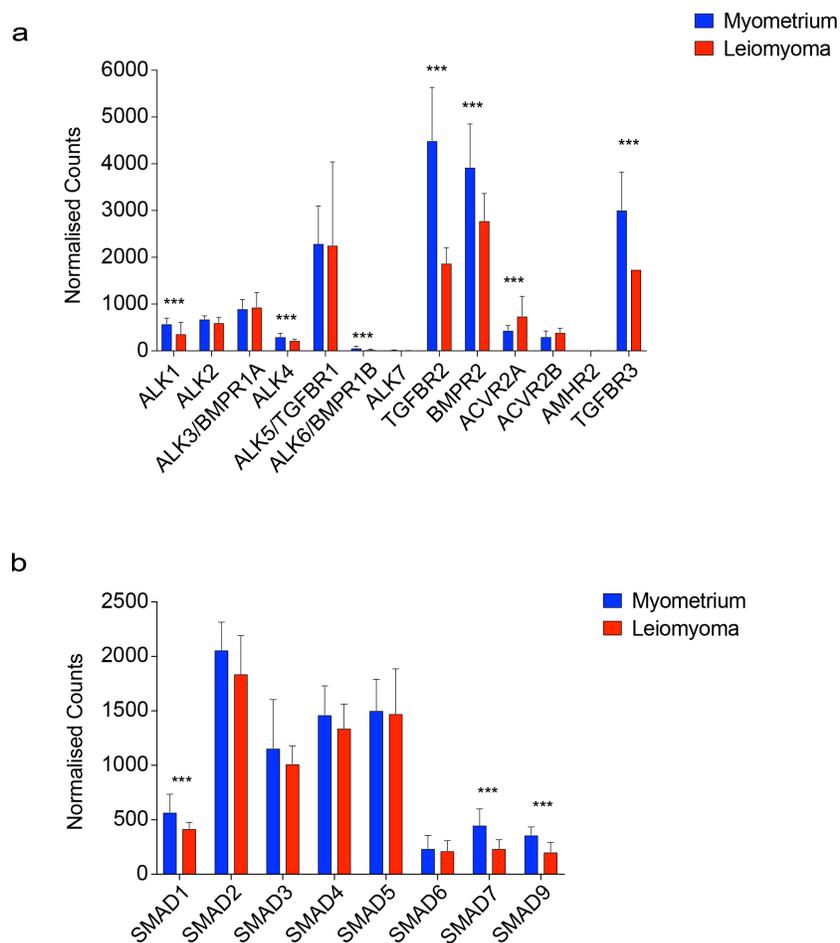
Figure 3.1.18. Transforming growth factor-β receptor and SMAD gene expression in myometrium and leiomyoma tissue. (a) Bar graph of TGF-β receptor gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (*n* = 15 patients). **(b)** Bar graph of SMAD gene expression in myometrium (blue bars) and leiomyoma (red bars) tissue samples (*n* = 15 patients). *** Indicates genes found to be differentially expressed at an FDR < 0.01.
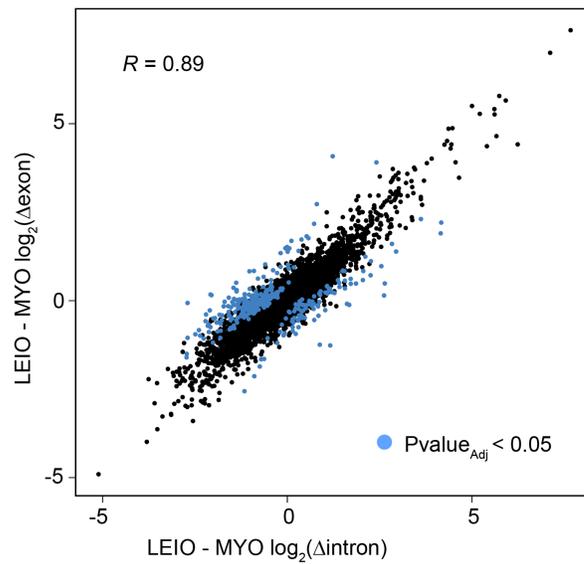
Figure 3.1.19. Posttranscriptional gene regulation in uterine leiomyomas.
Scatter plot of exon-intron split analysis (EISA) comparing changes in exonic
reads vs. changes in intronic reads of all expressed genes between
myometrium and leiomyoma tissues. *R* represents the Pearson correlation
coefficient.

TGF-β signalling (FDR < 0.05) [1,396-398]. In addition to microRNA-mediated posttranscriptional regulation, analysis of gene transcripts in uterine tissue samples revealed 699 genes that demonstrated usage of alternative transcripts, which includes alternative transcriptional starts sites as well as differences in exon usage, in leiomyomas compared to myometrium (Fig. 3.1.20). A significant number of these genes were ECM and ECM-associated genes (FDR < 0.01), indicating that alternative transcripts of ECM genes may be performing pro-fibrotic roles in uterine leiomyomas.

3.2. Histone acetylation changes occur disproportionately at intergenic regions in uterine fibroids

Acetylation of the histone H3 tail at lysine 27 (H3K27) is a post-translational modification that is highly correlated with active transcription at promoters and enhancers [321]. To provide epigenetic mechanistic insight into the predominantly transcriptional changes in gene expression observed by leiomyoma transcriptome profiling, chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) of H3K27Ac was performed (Table 3.1). High resolution mapping of histones on DNA was performed by enzymatic digestion of myometrium and leiomyoma tissue samples with micrococcal nuclease (MNase), an enzyme which initially endonucleolytically cleaves DNA in a sequence-preferential manner followed by exonucleolytic degradation of the DNA until the enzyme encounters a structural impediment to digestion such as histones or transcription factors and cofactors [399,400]. MNase digestion of chromatin resulted in a mono, -di, and -tri nucleosomal DNA profile (Fig. 3.2.1).
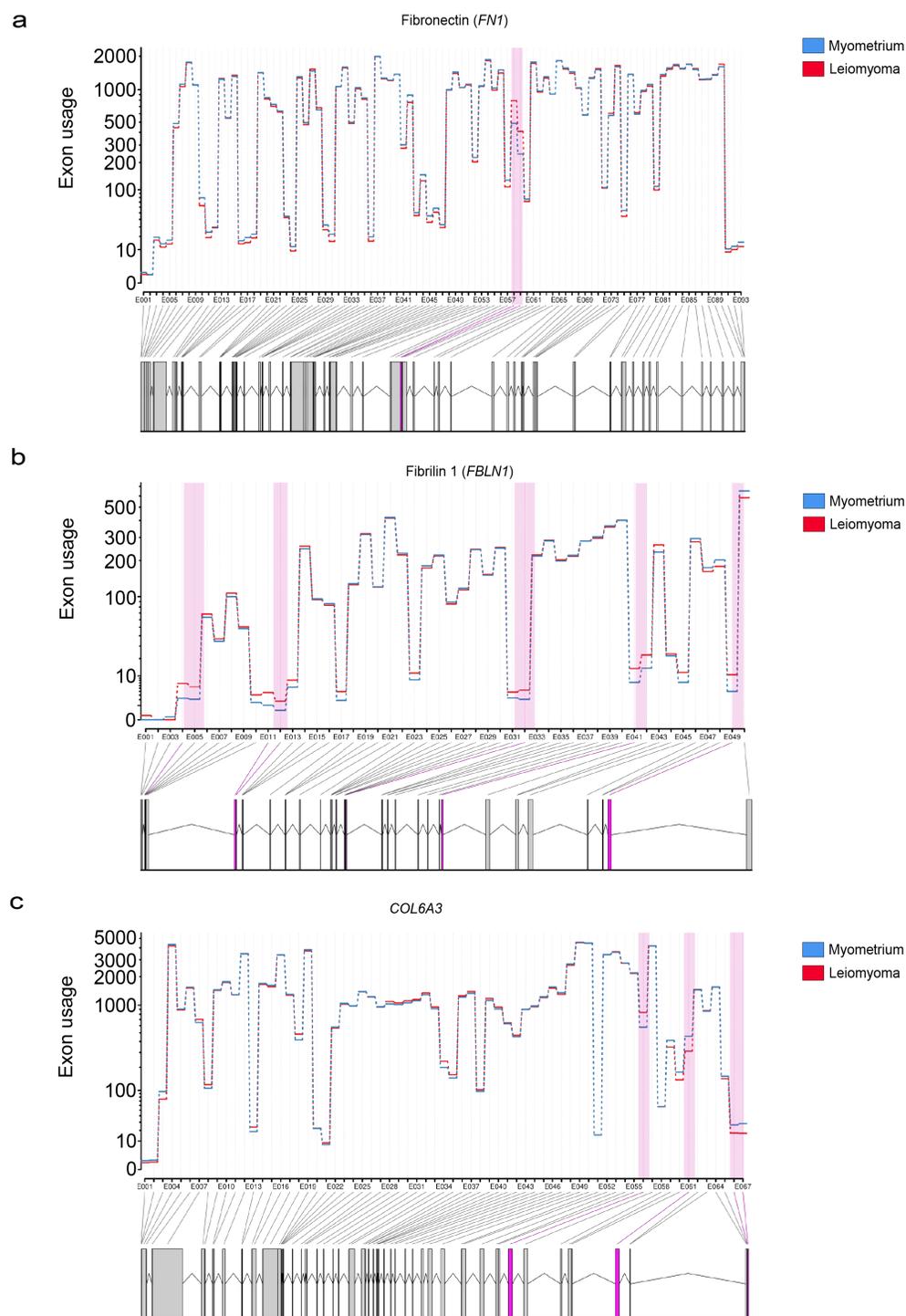
Figure 3.1.20. Differential exon usage in uterine leiomyomas. (a, b, c) Plot of normalised RNA reads for **(a)** *FN1,* **(b)** *FBLN1,* and **(c)** *COL6A3* gene transcripts identified as showing differential exon usage (FDR < 0.01).

IP	Total # of Peaks	# of differential Peaks
H3K27Ac	54,488	16,752
RNAPII	34,840	6,163
JUN	18,283	2,919
FOS	25,736	2,993
CDK8	44,402	4,696
MED12	52,033	6,857
CDK8 submodule (CDK8-MED12 overlap)	30,697	5,409

Table 3.1. Total ChIP consensus peaks identified from all myometrium and leiomyoma biological replicates. Consensus peaks for H3K27Ac, RNAPII, JUN, FOS, CDK8 and MED12 are shown. Numbers of differential peaks for each IP are noted. Total consensus peaks and differential peaks that are co-bound by CDK8 and MED12 (CDK8 submodule) are also listed.
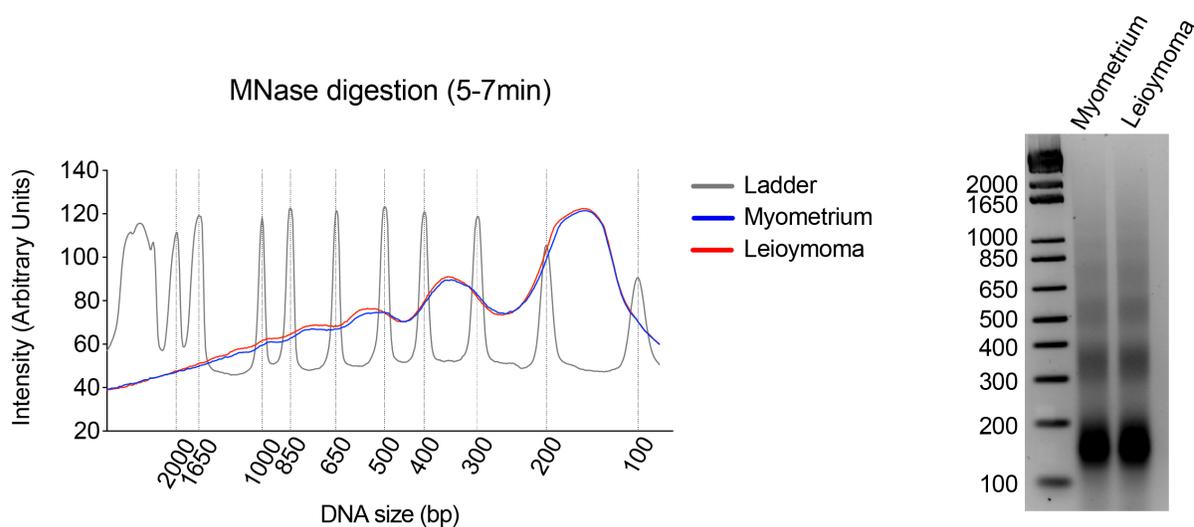
Figure 3.2.1. Myometrium and leiomyoma tissue MNase digestion profiles. Samples were digested for 5-7 minutes with Micrococcal nuclease to produce mono, di-, and tri-nucleosomal DNA for chromatin immunoprecipitation. DNA electrophoresis gel (right panel) of myometrium and leiomyoma MNase-digested chromatin is shown. Left panel is ImageJ-quantified signal of DNA gel bands.

Similar to RNA-seq samples, biological replicates of H3K27 acetylation ChIP experiments are highly correlated and also cluster by tissue type (Fig. 3.2.2). Significantly, 30% (n = 16,752/54,488) of all identified H3K27Ac regions are differentially acetylated with > 2-fold change in signal, and approximately 10% (n = 5,840/54,488) exhibit > 4-fold change in H3K27Ac signal (Fig. 3.2.3a, b, Fig. 3.2.4).

Analysis of genomic loci at sites of altered H3K27Ac signal revealed changes at predominantly promoter-distal regions, as evidenced by a significant deviation from the expected H3K27Ac signal distribution across the genome (Fig. 3.2.5, Fig. 3.26, ($\chi^2(4)$ = 1205, p < 0.05). While gene promoter acetylation signal accounts for approximately 17% of all H3K27 acetylated regions in myometrium and leiomyoma, acetylated promoters account for only 8% of all differentially acetylated regions. Conversely, higher than expected numbers of promoter-distal sites were differentially acetylated, with 85% of all differentially acetylated regions occurring at intergenic and intronic regions compared to 75% of sites in myometrium and leiomyoma. This suggests alterations at *cis* regulatory elements such as enhancers in uterine leiomyomas.

Interestingly, whilst analysis of dysregulated gene promoters shows that H3K27Ac signal changes are positively correlated with differential gene expression, a striking number of gene promoters were observed to have small changes in promoter H3K27 acetylation, with 46% of all dysregulated gene promoters shown to have little to no change in promoter H3K27 acetylation in uterine leiomyomas (Fig. 3.27). Identified examples of genes which, despite being overexpressed in leiomyomas, do not show a significant change in promoter H3K27 acetylation include Fibronectin 1 (*FN1*), a disintegrin and metalloprotease domain-containing protein 19 (*ADAM19*),
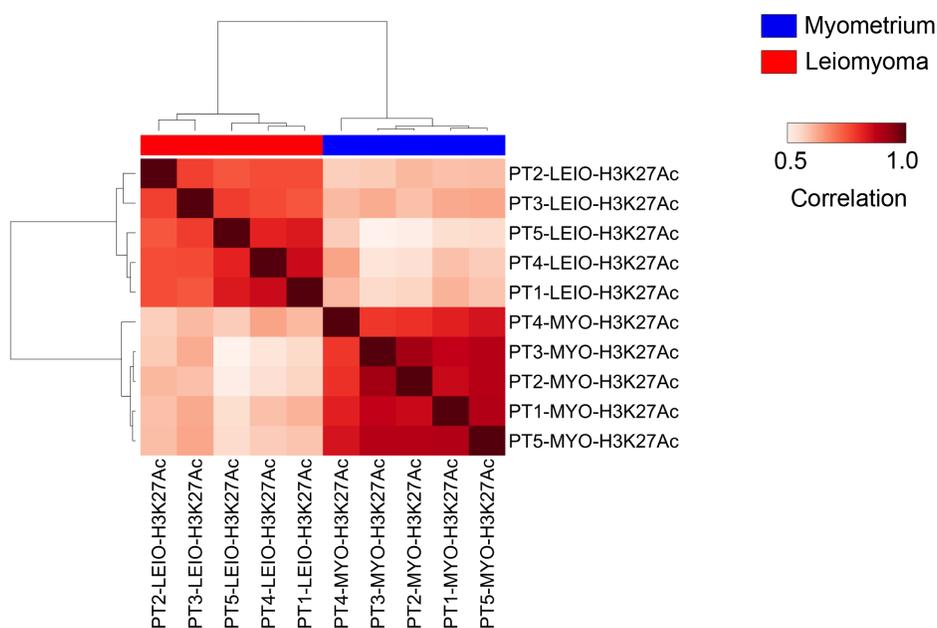
Figure 3.2.2. Correlation (Pearson) heat map of H3K27Ac ChIP affinity scores. Myometrium (blue) and leiomyoma (red) ChIP-seq read counts were used to determine sample affinity scores.
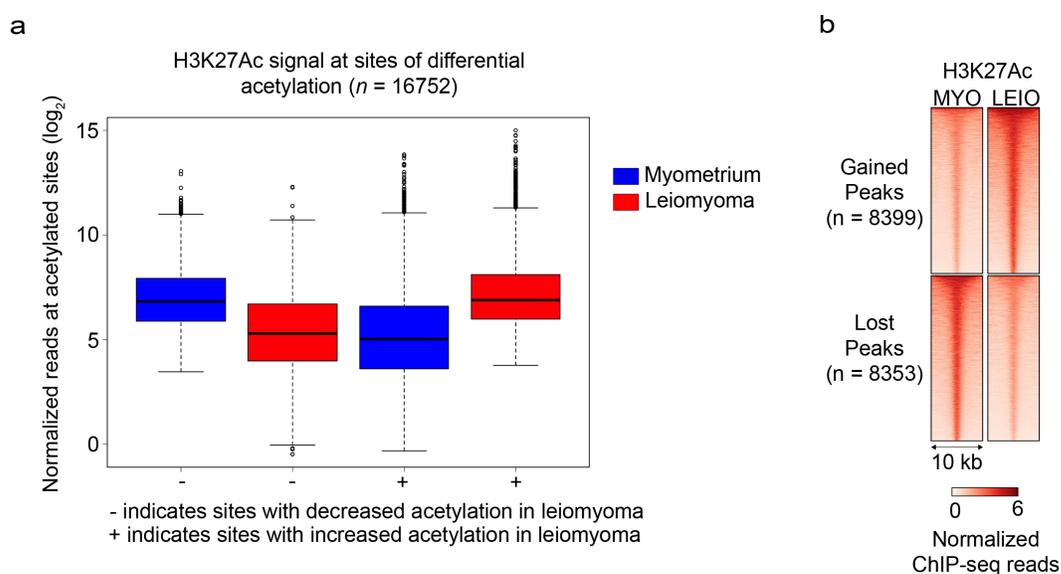
Figure 3.2.3. Differential H3K27 acetylation in uterine leiomyomas. (**a**) Box and whisker plot of normalised reads at differentially acetylated sites in myometrium and leiomyoma (left panel). Plots of sites with decreased H3K27Ac signal (-) and increased signal (+) are shown separately. Whiskers are 1.5 x interquartile range (IQR). (**b**) Heat map of normalised H3K27Ac ChIP-seq reads at differentially acetylated regions in myometrium vs. leiomyoma. Heat map represents average signal of 5 biological replicates.
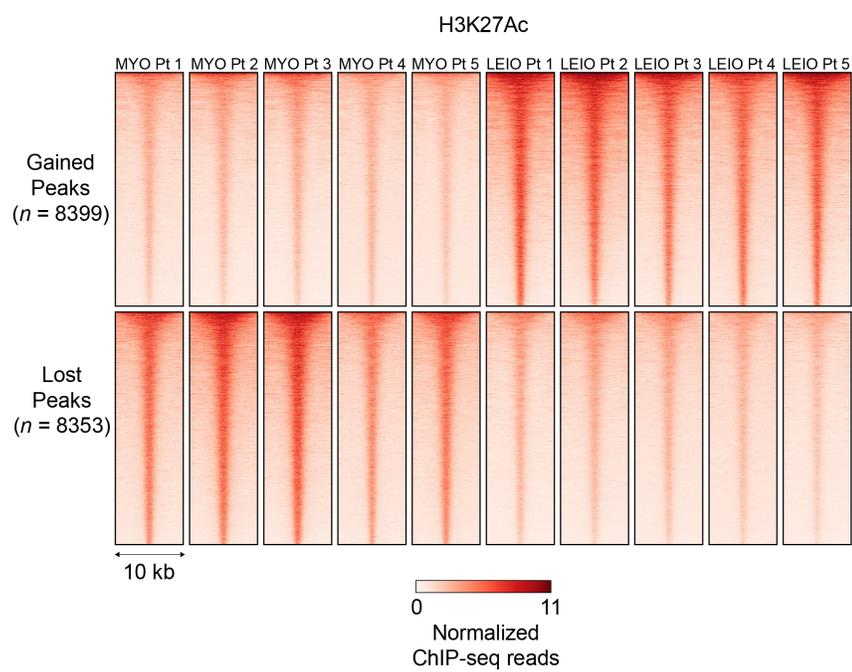
Figure 3.2.4. Heat map of H3K27Ac ChIP-seq reads at differentially acetylated regions in myometrium vs. leiomyoma. Normalised signal for each biological replicate is shown.
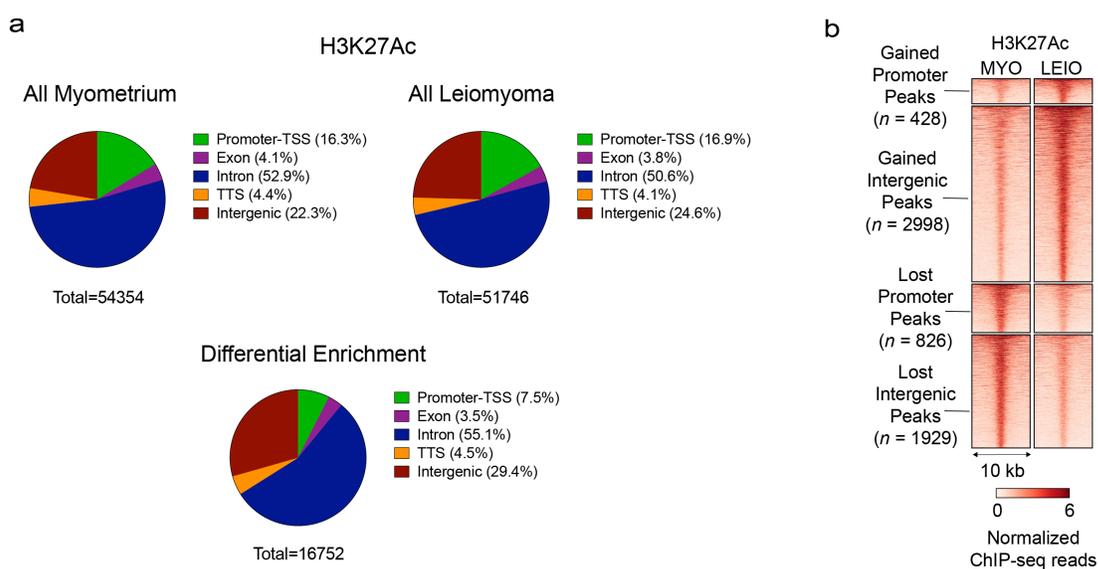
Fig. 3.2.5. Histone acetylation changes occur disproportionately at intergenic regions in uterine leiomyomas. (a) Pie charts of H3K27Ac signal distribution at genomic loci in myometrium (top left of left panel), leiomyoma (top right of left panel) and at differentially acetylated regions (bottom centre of left panel). **(b)** Heat map of normalised H3K27Ac ChIP-seq reads at gene promoters and intergenic loci with differential H3K27Ac signal. Heat map represents average signal of 5 biological replicates.
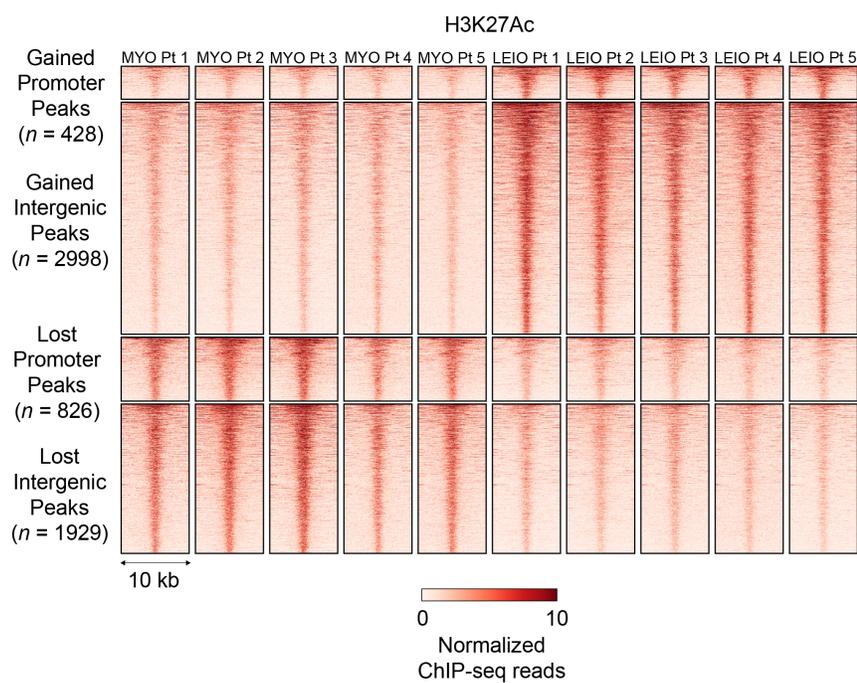
Fig. 3.2.6. Heat map of H3K27Ac ChIP-seq reads at gene promoters and intergenic loci with differential H3K27Ac signal. Normalised signal for each biological replicate is shown.
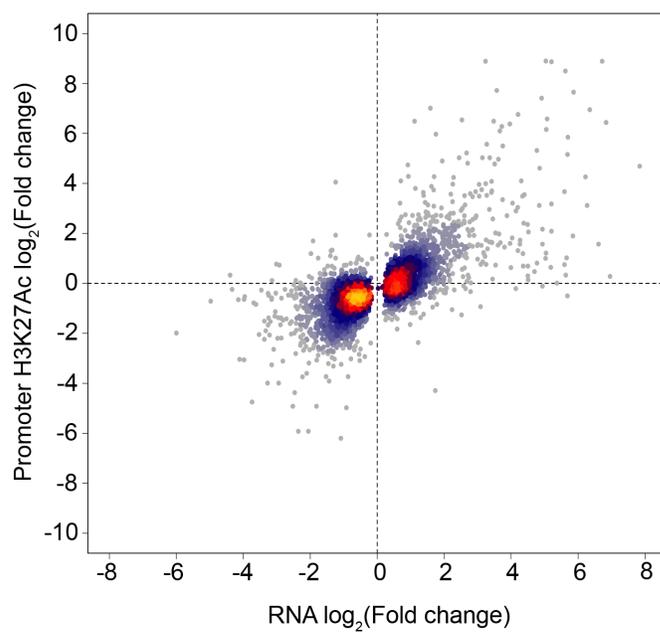
Figure 3.2.7. Correlation between differential gene expression and promoter H3K27Ac signal in uterine leiomyomas. Heat scatter plot of fold change in H3K27Ac signal against fold change in gene expression at promoters of differentially expressed genes ($n = 3,559$).
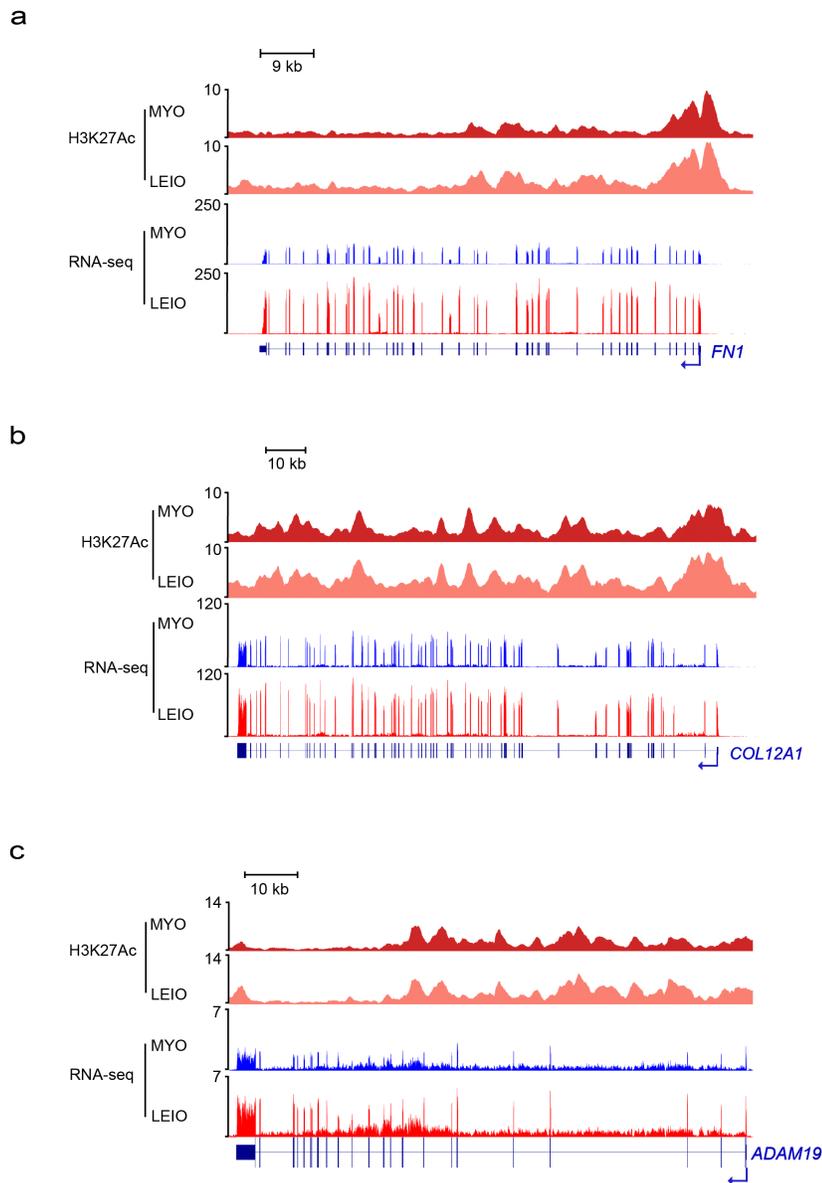
Fig. 3.2.8. Loci of dysregulated genes with unchanged promoter H3K27Ac signal. (**a, b, c**) Unchanged H3K27Ac signal in myometrium (red) and leiomyoma (salmon) at (**a**) *FN1*, (**b**) *COL12A1*, and (**c**) *ADAM19* genes. Normalised RNA reads for myometrium (blue) and leiomyoma (red) are also shown.

and collagen gene *COL12A1*, key genes involved in extracellular matrix formation and degradation (Fig. 3.2.8). In cases such as these, changes in enhancer activity, in addition to promoter acetylation changes, may together better explain the observed changes in gene expression. Given the large proportion of H3K27Ac cistrome that is modified in uterine leiomyomas and the significant number of promoters of differentially expressed genes with little to no change in H3K27 acetylation, it is predicted that modifications in enhancer regions play a role in uterine leiomyoma gene dysregulation.

3.3. Enhancer malfunction drives gene expression changes in uterine leiomyomas

H3K27Ac ChIP-seq data suggests enhancer dysfunction as a likely defining feature of leiomyoma transcriptional dysregulation. Assignment of enhancers to target gene promoters using nearest neighbour approaches has been shown to suffer from low accuracy. Therefore, to assign dysregulated enhancers to target genes and determine if they are directly linked to differential gene expression in uterine fibroids, promoter capture Hi-C was performed [363,401]. Interaction confidence scores for promoter interacting regions were calculated using the Capture Hi-C Analysis of Genomic Organisation (CHiCAGO) pipeline [390]. A stringent CHiCAGO-assigned contact score ≥ 5 has been demonstrated to be quite robust in identifying significant interactions. A reciprocal promoter capture Hi-C study of 949 promoter interacting regions demonstrated that over 90% of all high confidence interactions (i.e. CHiCAGO score ≥ 5) fell well within the *sdef* – determined consistency range between capture Hi-C and reciprocal capture Hi-C [402,403]. This demonstrates that a score ≥ 5 performs very well at identifying true positives whilst minimising the rate of false positives. In total, 163,712 enhancer-promoter contacts and 27,392 promoter-promoter contacts were identified with a contact score ≥ 5 in myometrium and/

or leiomyoma tissue samples (Table 3.2). This constitutes approximately 0.05% of all CHi-C identified contacts. The majority of promoter interacting CREs were located approximately 100-400kb from their respective promoters, with over 99% of contacts being made between genomic loci less than 1Mb apart (Fig. 3.3.1a). There was very little difference in distances made between two promoters (P-P) and contacts made between enhancers and promoters (E-P) (Fig. 3.3.1b).

Sub-threshold CHiCAGO scores between 3 and 5 ($3 \leq$ score < 5) may have a higher false positive rate but also identify meaningful interactions. 299,808 enhancer-promoter contacts and 48,766 promoter-promoter contacts were identified with a score between 3 and 5 in myometrium and/ or leiomyoma tissue samples. The high number of interactions recovered with a score between 3 and 5 may suggest that while a cut-off score ≥ 5 captures a significant number of high confidence interactions, many valid interactions may also be lost due to the high stringency of the interaction calls. Approximately 65% (14,289/22,101) of all assayed baits made at least one high confidence interaction with a distal region, with an average of 11.4 high confidence contacts per bait (Fig. 3.3.2a). This supports conclusions from previous studies that promoters receive multiple signal cues from different enhancers that may be bound by different transcription factors and complexes, resulting in an integrated transcriptional signal at gene promoters. However, individual distal regions with at least one high confidence interaction made a small number of contacts with promoters, with an average of 1.65 high confidence enhancer-promoter interactions per distal region, suggesting that distal contacts are bait-specific and do not function as signalling hubs that relay similar signal cues to multiple gene promoters (Fig. 3.3.2b).

Identified high confidence promoter contacts map to 98,957 unique promoter-distal regions in the human genome as determined by *Hin*dIII digestion of the genome (Fig. 3.3.3a).

	Total # of contacts	# of altered contacts
Enhancer-Promoter (E-P)	163,712	8,078
Promoter-Promoter (P-P)	27,392	1,428

Table 3.2. Promoter capture Hi-C identified contacts identified from all myometrium and leiomyoma biological replicates. Enhancer-promoter (E-P) and Promoter-promoter (P-P) contacts are listed, with altered contacts also divided into E-P and P-P contact groups.

a

b
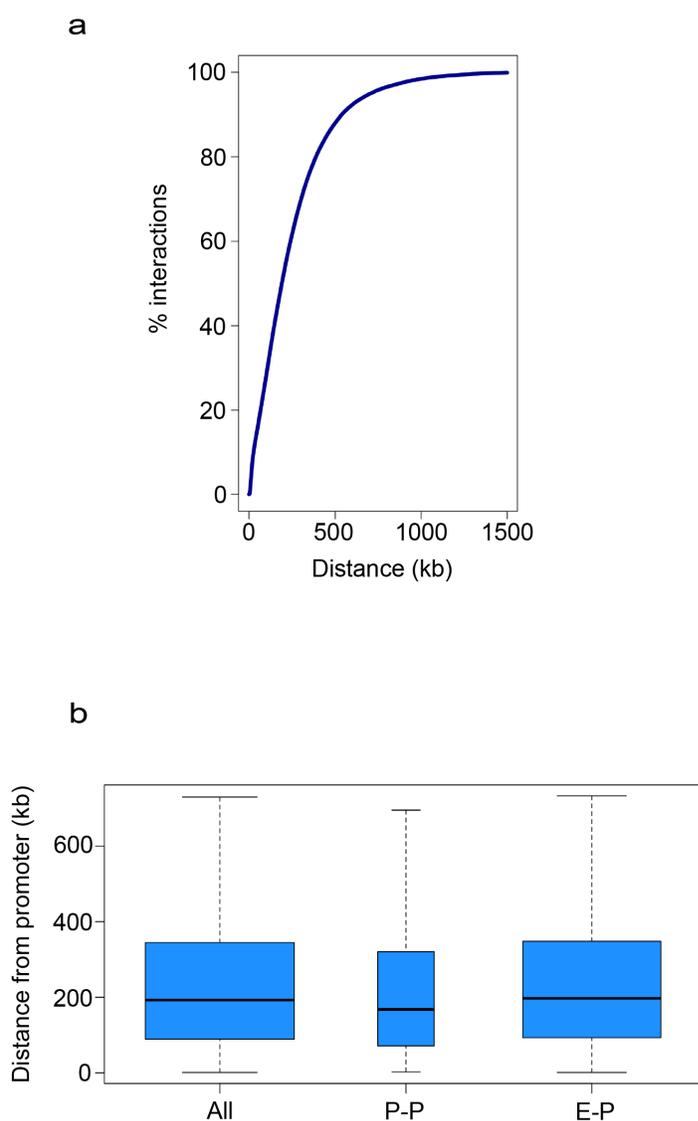
Fig. 3.3.1. Distance profiles of CHiCAGO-determined contacts. (a) Cumulative distribution of distances between promoters and their distal promoter interacting CREs. **(b)** Boxplot of distances for all contacts, promoter-promoter (P-P) contacts, and enhancer-promoter contacts.
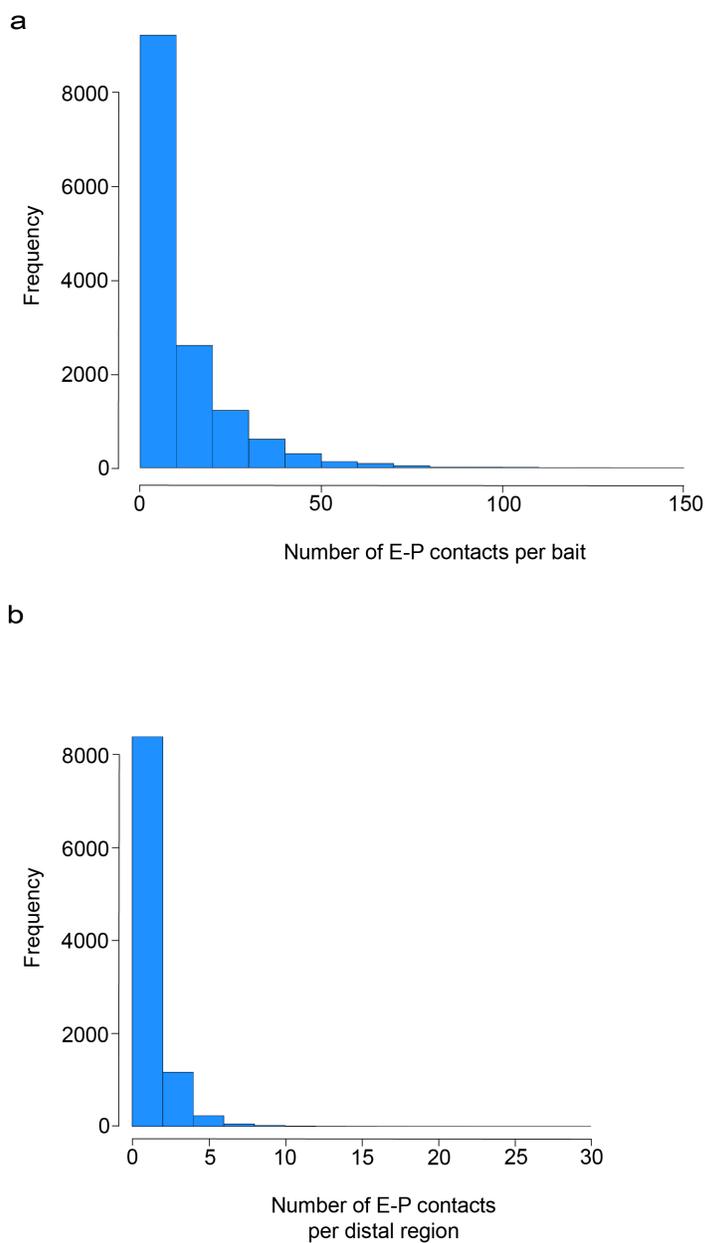
Fig. 3.3.2. Enhancer-promoter contact frequency distributions. (**a**) Histogram of high confidence contacts made by individual promoters and all of their corresponding CREs i.e. viewpoint is from gene promoters. (**b**) Histogram of high confidence contacts made by individual CREs with their corresponding promoters i.e. viewpoint is from CREs.

a

HindIII fragments
overlapping H3K27Ac

HindIII fragments overlapping
CHi-C interactions

77552   37822   61135

b
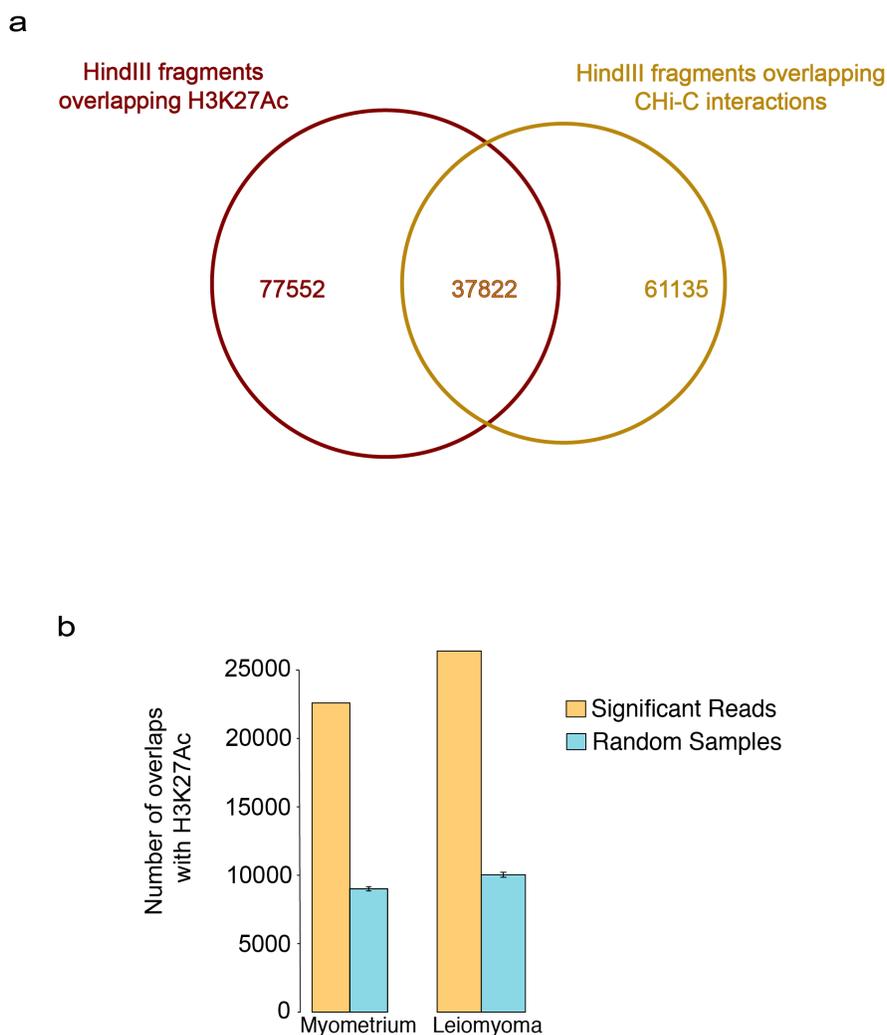
Fig. 3.3.3. Promoter contacts overlap with H3K27 acetylated regions. (a) Venn diagram of promoter contacts overlapping all acetylated H3K27 regions in myometrium and leiomyoma tissue. *Hin*dIII digest fragments of the human genome are used to define unique, non-overlapping regions in the genome, to which H3K27Ac peaks and enhancer-promoter contacts are then assigned. See CHi-C data processing and analysis in methods. **(b)** Bar graph of H3K27Ac peak enrichment in regions containing enhancer-promoter contacts. Expected number of H3K27Ac peaks (Random Samples) that would overlap with enhancer-promoter contacts by chance (blue bars, 95% confidence interval) and the observed H3K27Ac peaks that overlap with enhancer-promoter contacts in myometrium and leiomyoma (orange bars) are shown.

Promoter contacts significantly overlapped with H3K27 acetylated regions, co-mapping to

37,822 unique promoter-distal regions (Fig. 3.3.3a, b). Importantly, promoter contacts

overlapped with differentially acetylated H3K27 enhancer regions, co-mapping to 10,769 unique

promoter-distal regions (Fig. 3.3.4). Using promoter capture Hi-C to unambiguously assign

enhancers to their corresponding promoters, we identified 2,715 enhancers highly enriched or

depleted of H3K27Ac signal that were directly linked with differentially expressed genes (Fig.

3.3.5, Fig. 3.3.6). 1,835 enhancers were associated with induced or repressed genes exhibiting

little to no change in promoter H3K27 acetylation. In addition, the magnitude of enhancer

acetylation signal changes was significantly greater than acetylation changes at dysregulated

gene promoters. These results suggest that in addition to changes at promoters, altered H3K27Ac

at enhancers plays a significant role in aberrant transcription of genes in uterine leiomyomas,

including genes with unchanged promoter-proximal H3K27Ac signal.

Capture Hi-C studies have previously demonstrated a signal dependent alteration of

enhancer-promoter contacts during differentiation [364,365]. We investigated whether promoter

contact strength is also altered in a disease dependent manner in *MED12* mutant uterine fibroids.

8,078 altered enhancer-promoter contacts were identified, with 1,974 contacts displaying > 2-

fold change in signal (Table 3.2). 25% ($n = 2,040/8,078$) of contacts were associated with

differentially expressed genes, which ontology analysis revealed to be enriched for genes

involved in extracellular matrix organisation (Fig. 3.3.7a, b). Interestingly, baits associated with

altered contacts made more contacts with distal regions (Fig. 3.3.8). Moreover, baits associated

with altered contacts and differential H3K27Ac signal made even more contacts with distal

regions. A survey of altered contacts overlapping enhancers with differential H3K27Ac signal
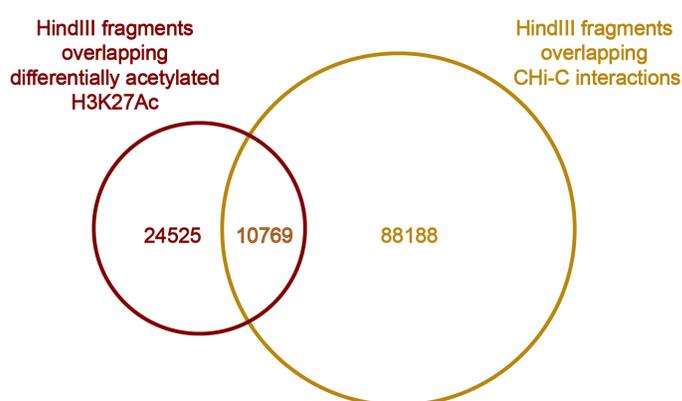
Fig. 3.3.4. Promoter contacts overlap with differentially acetylated regions.
Venn diagram of enhancer-promoter contacts overlapping differentially
acetylated H3K27 regions in myometrium and leiomyoma tissue. *Hin*dIII digest
fragments of the human genome are used to define unique, non-overlapping
regions in the genome, to which H3K27Ac peaks and enhancer-promoter
contacts are then assigned. See CHi-C data processing and analysis in methods.
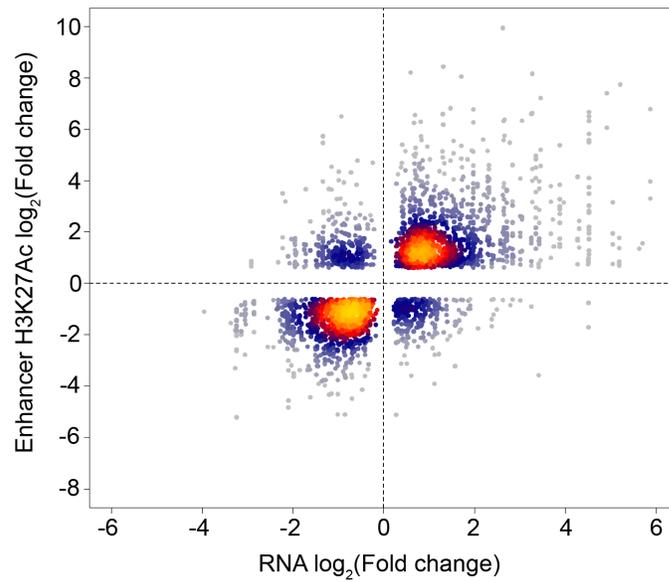
Fig. 3.3.5. Correlation between differential gene expression and enhancer H3K27Ac signal in uterine leiomyomas. Heat scatter plot of fold change in gene expression against fold change in H3K27Ac signal at enhancers associated with differentially expressed genes ($n = 2{,}715$ enhancer regions).
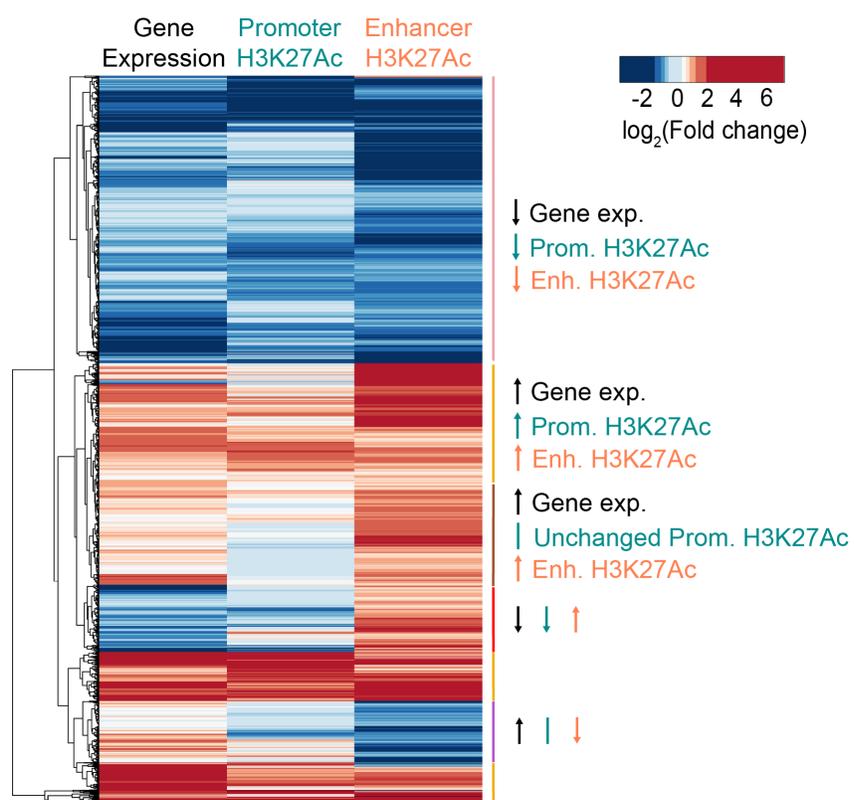
Fig. 3.3.6. Heat map of changes in gene expression, promoter, and enhancer H3K27Ac of differentially expressed genes. Genes are hierarchically clustered by similarity in profiles of differential gene expression and H3K27 acetylation ($n = 3{,}183$).
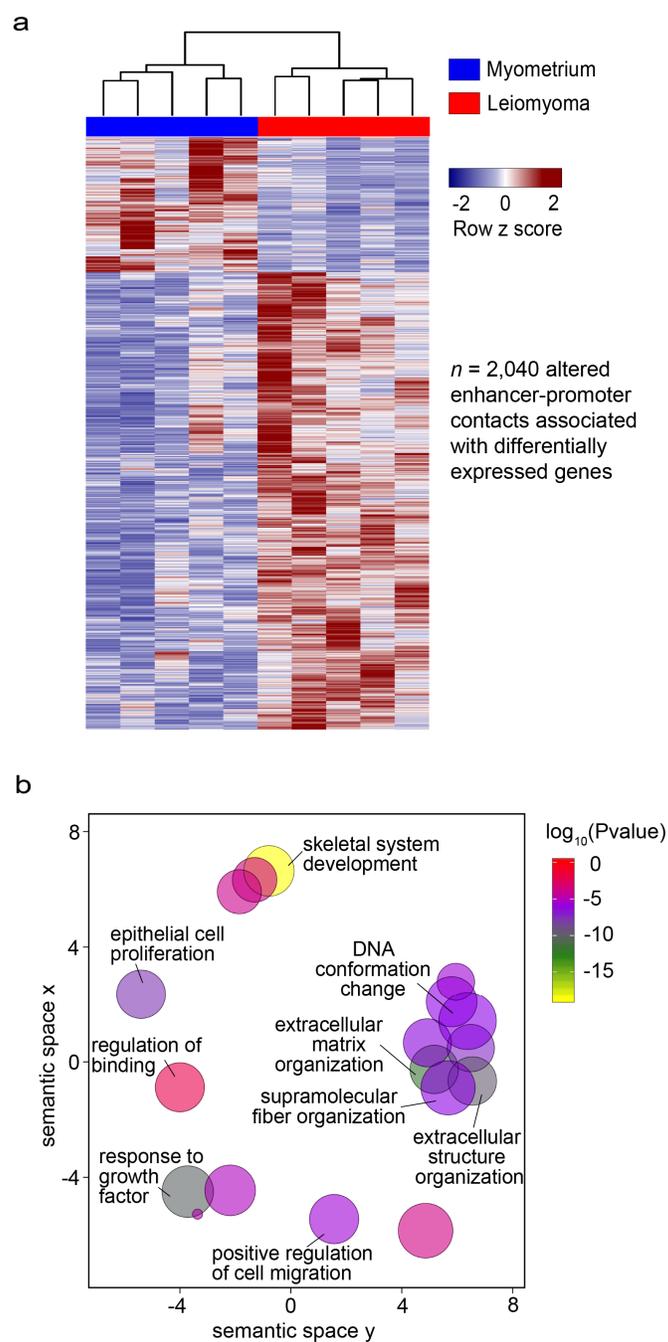
Fig. 3.3.7. Altered enhancer-promoter contacts in uterine leiomyomas. (a) Heat map of altered enhancer-promoter contacts associated with differentially expressed genes (n = 2,040 contacts). **(b)** Scatter plot of confidence scores for enriched gene ontologies associated with differentially expressed genes with altered enhancer-promoter contacts, with ontologies clustered by functional similarity in the semantic space.
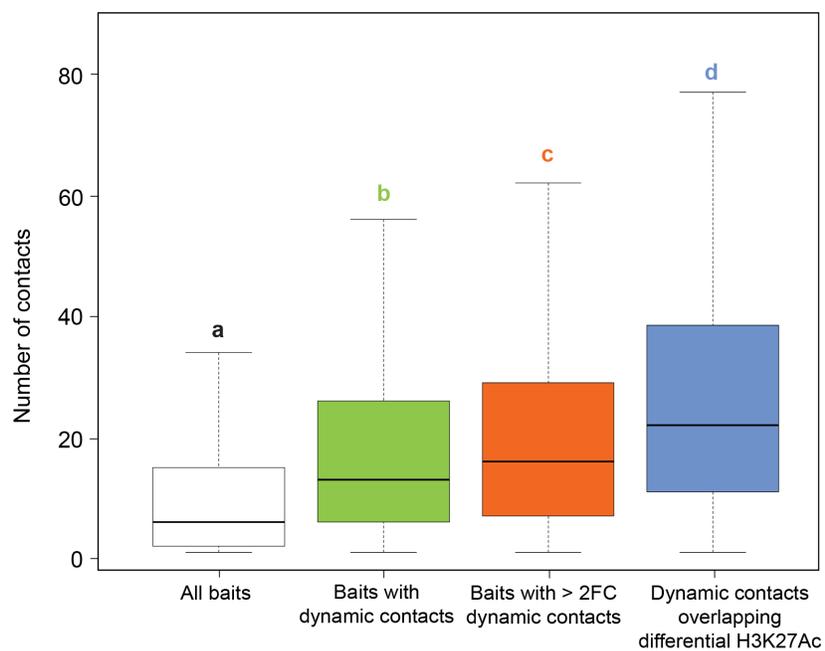
Fig. 3.3.8. Baits associated with altered enhancers make more contacts with promoter distal regions. Box and whisker plot of the number of high confidence contacts (i.e. score > 5) made by all baits, baits with dynamic contacts and baits with dynamic contacts that also overlap differentially acetylated H3K27 regions.

shows that regions with increased acetylation are associated with increased enhancer-promoter contact strength whereas regions depleted of H3K27Ac signal are linked to both contacts with increasing and decreasing strength (Fig 3.3.9a). Approximately 42% (521/1,229) of all altered contacts overlapping enhancers with differential H3K27Ac signal are associated with differentially expressed genes. Key extracellular matrix associated genes that demonstrated a change in gene expression accompanied by a change in enhancer-promoter contact strength include *FN1*, *ADAM19*, ETS proto-oncogene 2 (*ETS2*), as well as collagen genes *COL6A3*, and *COL12A1* (Fig. 3.3.10a, b, Fig. 3.3.11a, b, and Fig. 3.3.12a).

Depending on the status of H3K27 acetylation and promoter contact strength, modified promoter-distal CREs can be classified into three distinct categories: (i) regions with unaltered promoter contact strength and differential H3K27 acetylation; (ii) regions with altered promoter contact strength and unchanged H3K27 acetylation; and (iii) regions with altered promoter contact strength and differential H3K27 acetylation (Fig. 3.3.9b). Previous studies demonstrate that promoter contacts are cell type-defining features and as such, the majority of contacts are stable and unchanged [363,364,402]. However, despite the majority of contacts being stable, a subset are dynamic and do correlate with major events in the cell, such as changes in chromatin occupancy of transcription factors that have been demonstrated to be important in cell differentiation. Consistent with these previous studies, differentially acetylated enhancer regions with unaltered contact strength constitute the largest class of modified enhancers in leiomyomas, highlighting that enhancer architecture changes primarily consist of changes in histone modification signals. However, 27% of altered enhancer regions do show changes in promoter contact strength, which is a novel finding in a disease state. Although the significance of altered
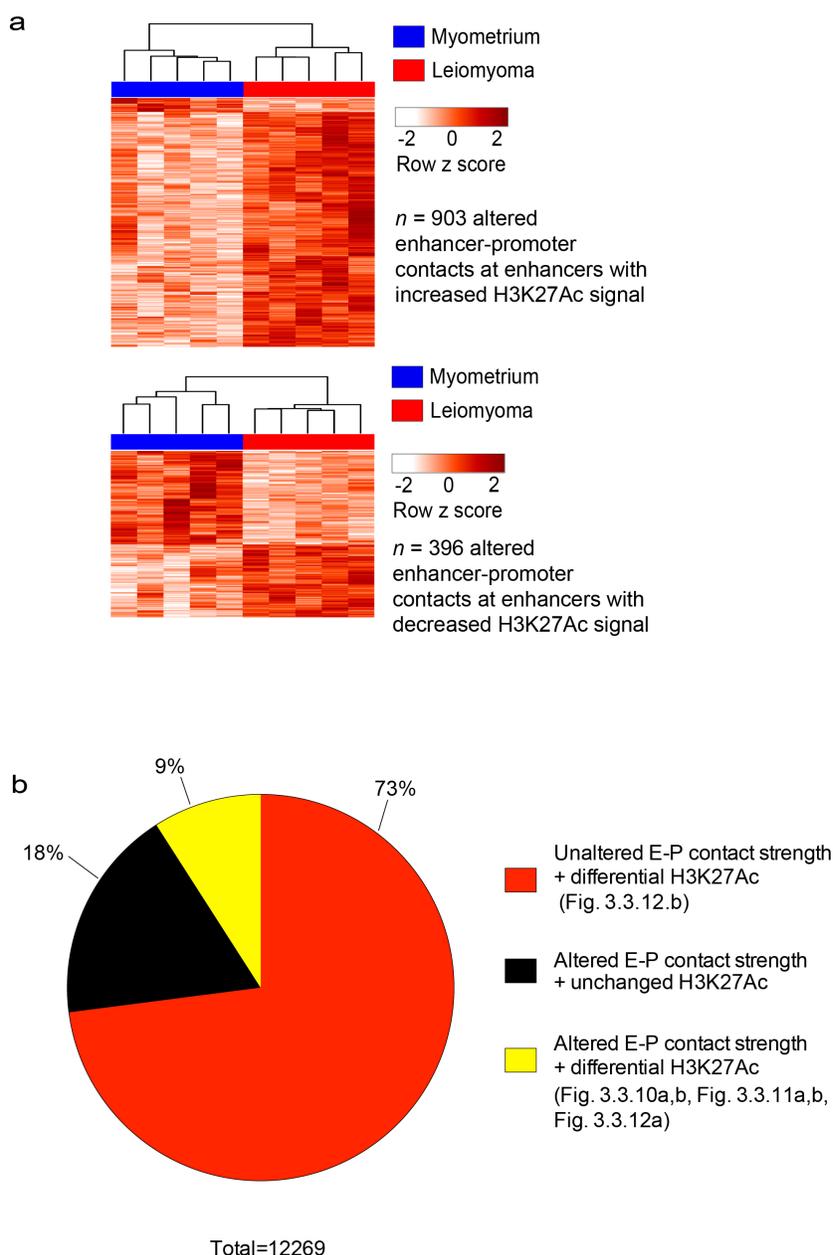
Fig. 3.3.9. Distinct categories of promoter-distal CREs. (**a**) Heat maps of altered enhancer-promoter contacts that overlap with increasing H3K27Ac (top panel, *n* = 903 contacts) and decreasing H3K27Ac (bottom panel, *n* = 396 contacts) at enhancer regions in myometrium and leiomyoma tissues. (**b**) Pie chart of modified cis-regulatory promoter-distal regions classified according to H3K27 acetylation status and alterations in promoter contact strength.
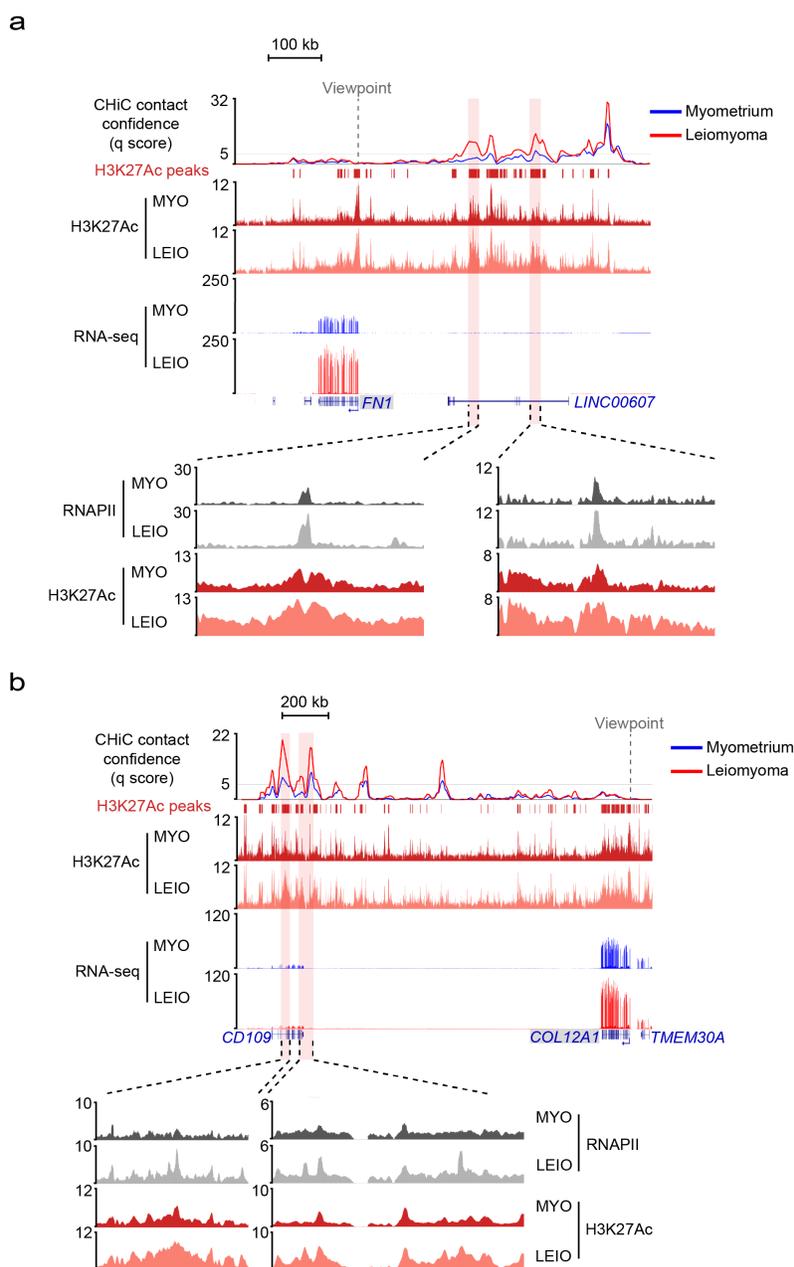
Fig. 3.3.10. Genomic loci associated with dysregulated genes that have unchanged promoter H3K27Ac signal. (a, b) Genomic loci for *FN1* (**a**) and *COL12A1* (**b**) genes that have unchanged H3K27Ac signal at promoters. Altered confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. In addition, H3K27Ac ChIP-seq and RNA-seq genomic tracks are shown, with H3K27Ac and RNAPII ChIP-seq signal in regions containing altered contacts highlighted (bottom zoomed in insert).
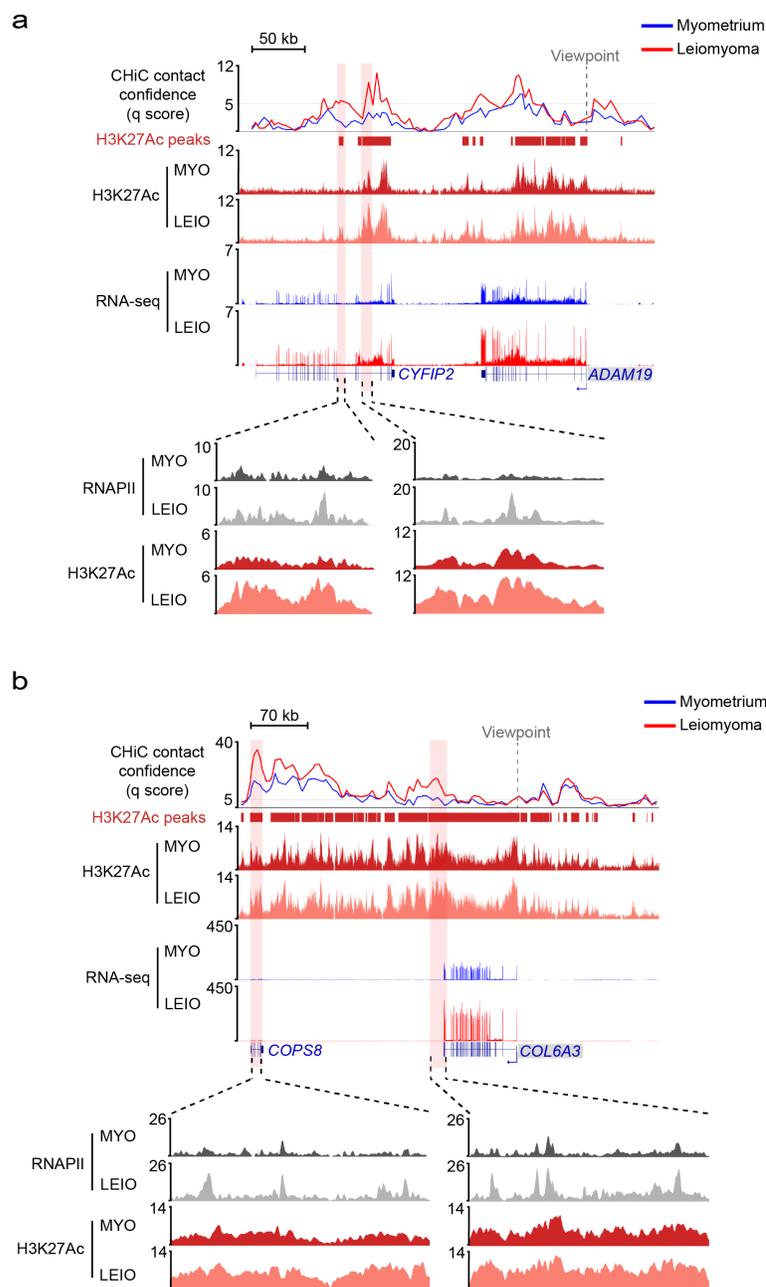
Fig. 3.3.11. Genomic loci associated with dysregulated genes that also have altered enhancer-promoter contacts. (**a, b**) Genomic loci for *ADAM19* (**a**) and *COL6A3* (**b**) genes, which are both associated with altered enhancer-promoter contacts. Altered confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. In addition, H3K27Ac ChIP-seq and RNA-seq genomic tracks are shown, with H3K27Ac and RNAPII ChIP-seq signal in regions containing altered contacts highlighted (bottom zoomed in insert).
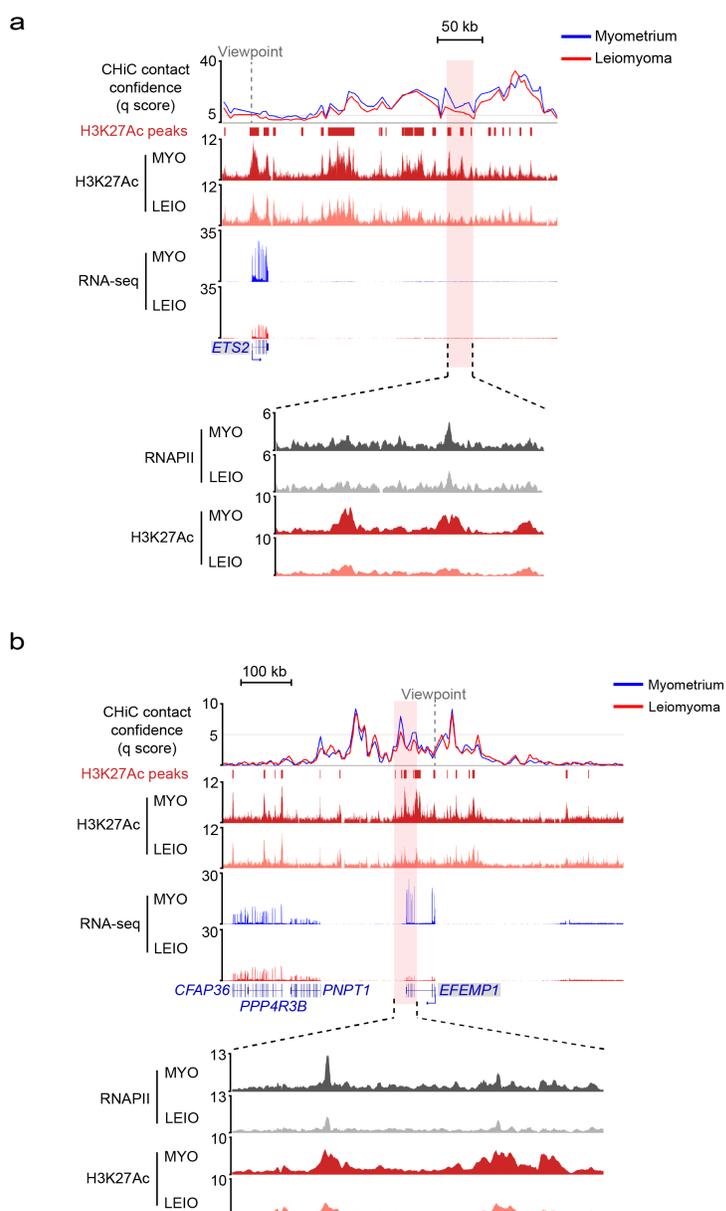
Fig. 3.3.12. Genomic loci containing promoter-distal CREs belonging to different categories. (**a, b**) Genomic loci for *ETS2* (**a**) and *EFEMP1* (**b**) genes. Altered enhancer architecture involving changes to both H3K27Ac signal and enhancer-promoter contact strength (**a**) or changes in H3K27Ac signal at enhancers with stable, unchanging enhancer-promoter contact strength (**b**). Confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. In addition, H3K27Ac ChIP-seq and RNA-seq genomic tracks are shown, with H3K27Ac and RNAPII ChIP-seq signal in regions containing altered contacts highlighted (bottom zoomed in insert).

contacts in uterine leiomyomas is unknown, the significance of altered contacts as demonstrated in cell differentiation may indicate a role for altered contacts in leiomyoma disease pathogenesis. Motif analysis of distal ends of enhancer-promoter contacts revealed significant CTCF binding motif enrichment, suggesting intra-TAD binding of CTCF and possible involvement in enhancer-promoter looping in uterine tissues (Fig. 3.3.13a). YY1 binding motifs were not identified in this analysis, although this may be due to the very low resolution obtained from motif analysis carried out on promoter capture Hi-C data of *Hin*dIII digested fragments. The average length of *Hin*dIII digested fragments is 3,590bp and over 50% of all fragments are shown to be greater than 2,000bp, with a median length of 2,215bp (Fig. 3.3.13b). As a result, other factors with smaller binding motifs than CTCF and that generally have a focused peak width of 100-300bp may be missed in the analysis.

Capture Hi-C also revealed instances where decreased H3K27Ac signal in leiomyoma samples at an enhancer associated with a gene occurred concomitantly with an increase in H3K27Ac signal at another enhancer associated with the same gene. 662 genes were shown to exhibit cases of differential enhancer usage, involving a total of 3,884 enhancers. Differential enhancer usage was seen in both up and down-regulated genes and occurred primarily at enhancers with stable, unaltered promoter contacts, although a few cases were also observed in a subset of genes that have enhancers with altered contact strength (Fig. 3.3.14a, b).

Promoter capture Hi-C performed on myometrium and leiomyoma tissue samples demonstrates multiple changes in enhancer architecture in uterine leiomyomas. This includes aberrant enhancer acetylation, which is correlated with changing gene expression for genes where promoter acetylation is unchanged as well as for genes with differential promoter

a

TF motif enrichment at CREs contacting promoters

Motif family	% sites	P value	Motif
CTCF	13.93%	1e-202	
CTCFL	17.18%	1e-77	
ZNF16	1.91%	1e-14	

b
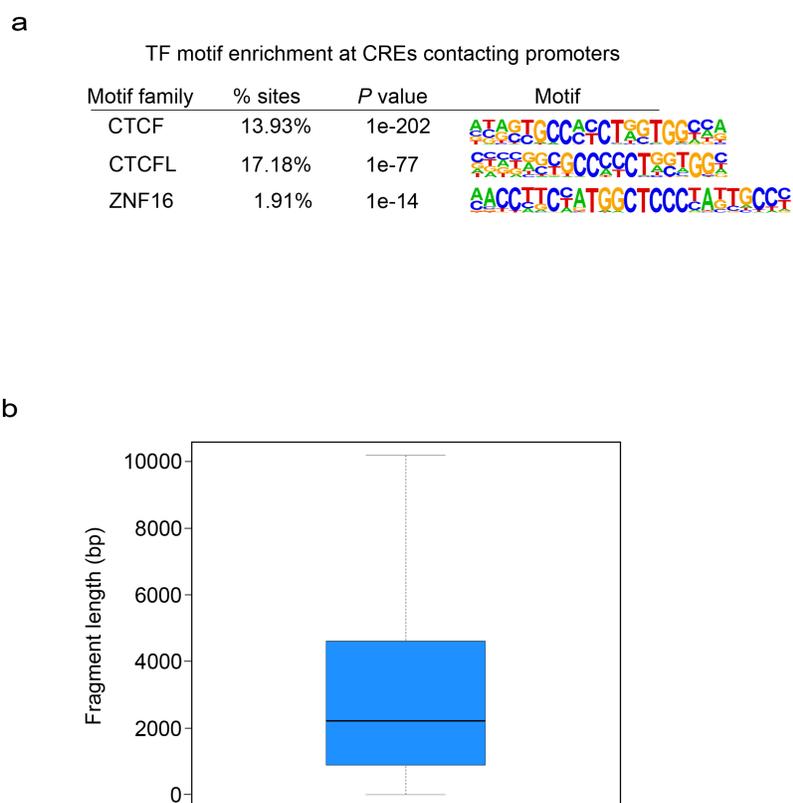
Fig. 3.3.13. CTCF intra-TAD motif enrichment at distal CREs. (a) Top identified transcription factor motifs (*P* value ranking) enriched in distal *cis* regulatory elements that make contacts with promoters and have at least one contact with a contact score > 5. **(b)** Box and whisker plot of *Hin*dIII fragment lengths for the whole human genome.
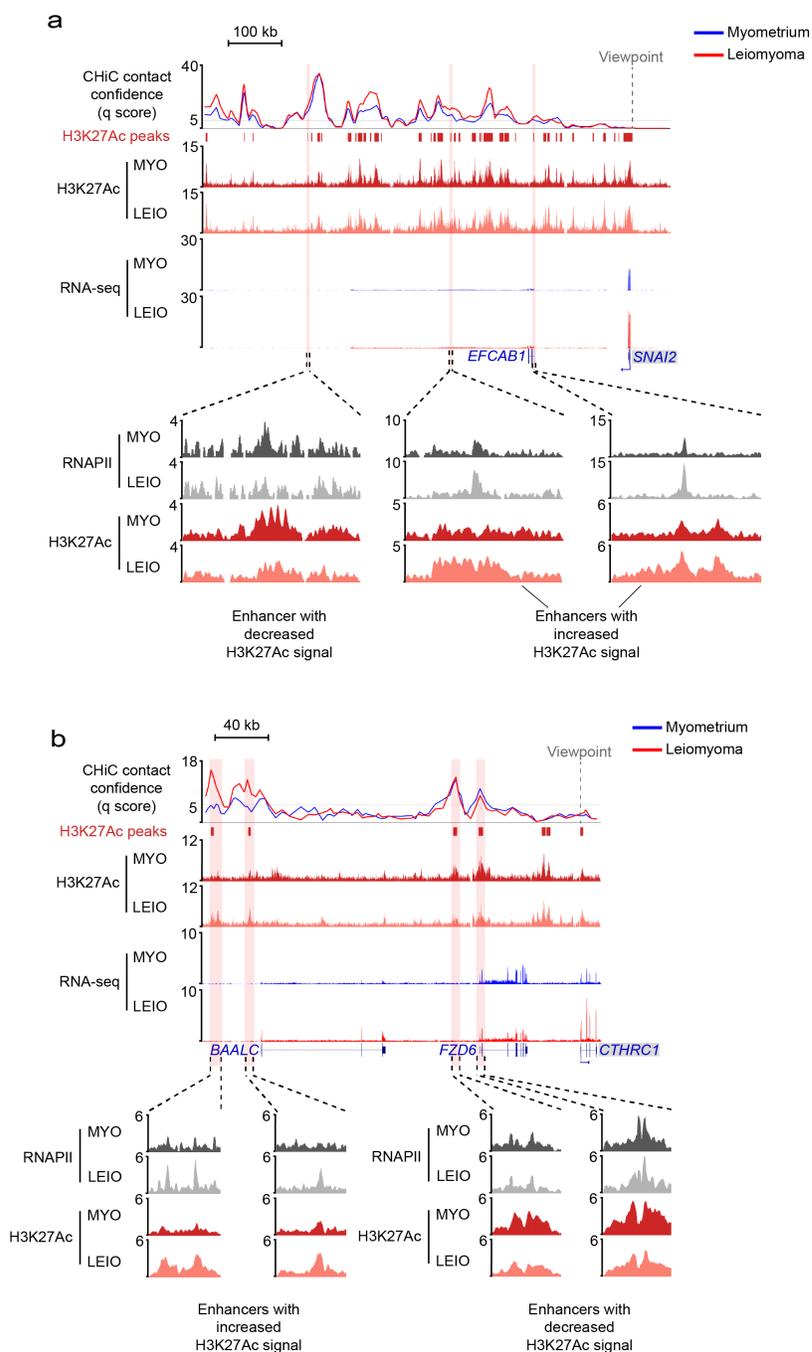
Fig. 3.3.14. Differential enhancer usage in uterine leiomyomas. (**a, b**) Genomic loci for *SNAI2* (**a**) and *CTHRC1* (**b**), both of which exhibit cases of differential enhancer usage. Differential enhancer usage generally involves changes in distal regions with stable enhancer-promoter contacts and differential H3K27Ac signal (**a**) but may involve enhancers with altered enhancer-promoter strength as well as differential H3K27Ac signal (**b**).

acetylation. This highlights enhancer malfunction, a previously undescribed feature in uterine leiomyoma pathobiology, as an important mechanism of leiomyoma gene dysregulation and in the case of differentially expressed genes with stable promoter H3K27Ac signal, provides a mechanism of gene dysregulation in the absence of promoter-proximal chromatin changes. The observation of altered contacts in leiomyomas also suggests a partial re-wiring of chromatin architecture in uterine leiomyomas.

3.4. Activator protein 1 (AP-1) transcription factor complex expression and chromatin occupancy is perturbed in uterine leiomyomas

Given the predominantly promoter-distal changes in H3K27 acetylation and the alterations in promoter contact strength, differentially acetylated enhancer regions were analysed for the presence of transcription factor DNA binding motifs that may be important in regulating gene transcription in leiomyomas. The AP-1 transcription factor motif was enriched at differentially acetylated enhancers (Fig. 3.4.1a). Analysis of transcription factor motifs at promoters of differentially expressed genes also showed an enrichment of the AP-1 binding motif, suggesting AP-1 driven dysregulation of direct AP-1 target genes (Fig. 3.4.1b, FDR < 0.05). Notably, expression of the *JUN*, *FOS*, and *ATF* families of genes was down regulated in leiomyoma tissue (Fig. 3.4.2). This is consistent with previous studies highlighting greater than five-fold reduction in *FOS* and *JUN* mRNA levels in uterine fibroids [355-358]. Interestingly, AP-1 subunit genes exhibit little to no change in promoter H3K27Ac. A survey of enhancers associated with *JUN* and *FOS* showed differential H3K27Ac signal at a subset of enhancers, with differential enhancer usage observed at enhancers contacting the *JUN* promoter, although enhancer-promoter contact strength was unaltered at both *JUN* and *FOS* enhancers (Fig. 3.4.3a, b).

a

TF motif enrichment at enhancer sites with differential H3K27 acetylation

Motif family	% sites	P value	Motif
Jun-AP1	31.54%	1e-58	
Fosl2	39.92%	1e-39	
Bach2	24.19%	1e-24	
NF-E2	8.36%	1e-21	

b
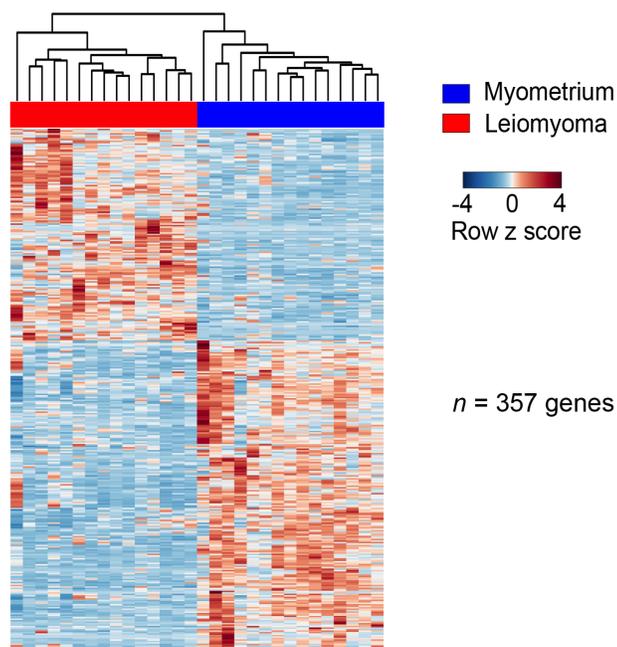
Fig. 3.4.1. AP-1 motif enrichment at differentially acetylated enhancer regions in uterine leiomyomas. (**a**) Top identified transcription factor motifs (*P* value ranking) enriched in enhancer regions with differential H3K27Ac signal between myometrium and leiomyoma. (**b**) Heat map of differentially expressed genes with a promoter-proximal AP-1 binding site. Gene expression levels relative to the mean expression are shown as row z scores.
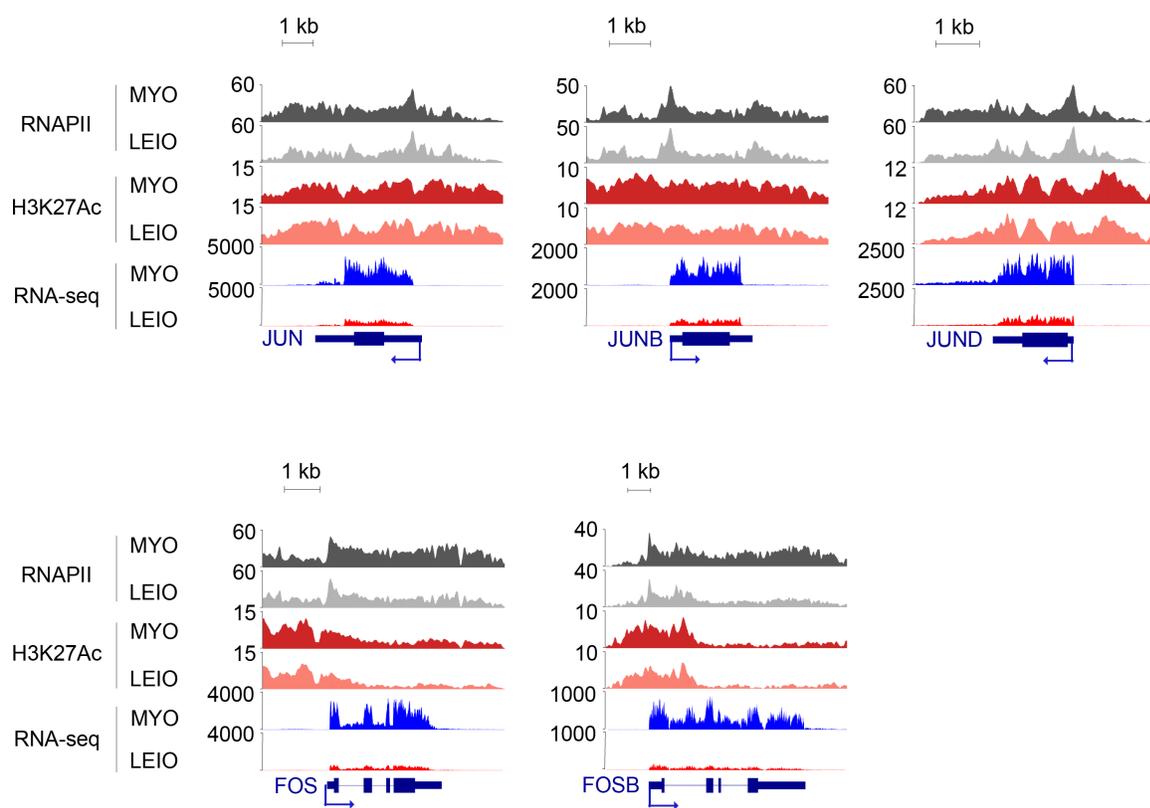
Fig. 3.4.2. AP-1 subunit gene loci. H3K27Ac and RNAPII ChIP-seq genomic tracks for AP-1 subunits *JUN*, *JUNB*, *JUND*, *FOS* and *FOSB*.
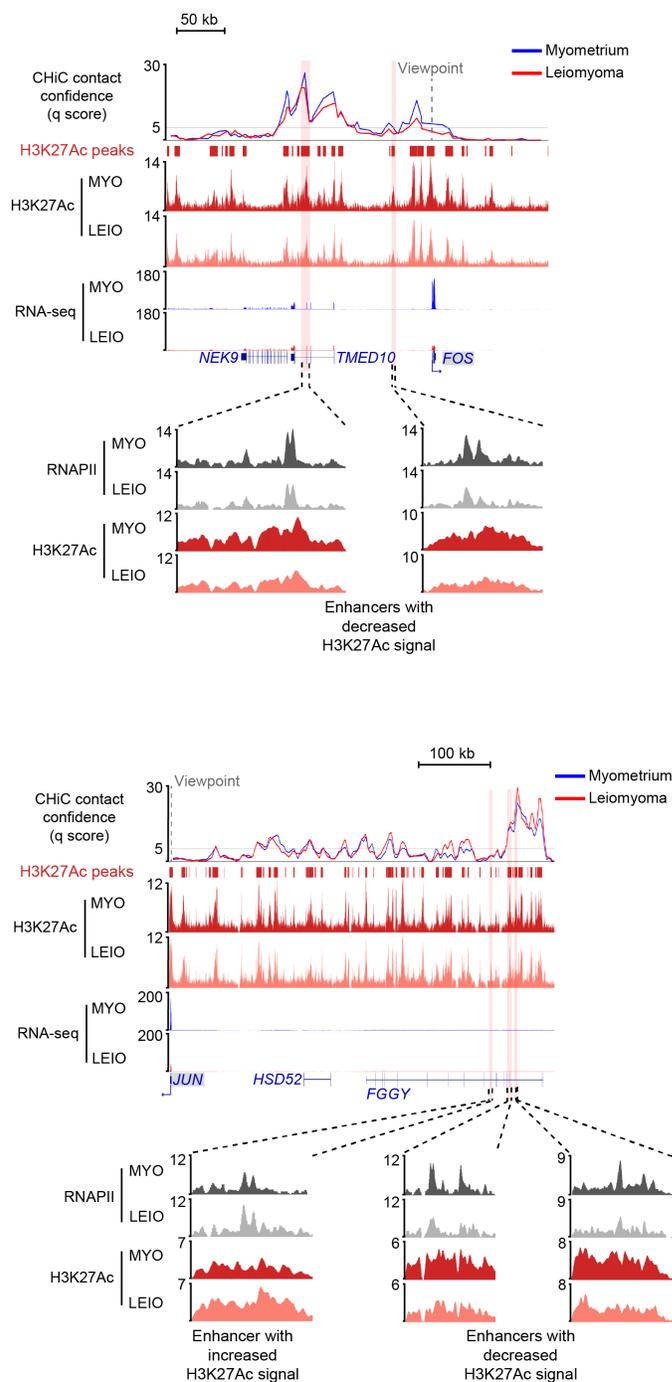
Fig. 3.4.3. Genomic loci associated with dysregulated AP-1 genes. (a, b) Genomic loci for *FOS* (**a**) and *JUN* (**b**) showing RNA expression and enhancer-promoter contacts in myometrium and leiomyoma. Confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. H3K27Ac and RNAPII ChIP-seq at sites of increased or decreased H3K27Ac signal are highlighted.

Whilst the mechanisms of AP-1 gene regulation may be varied, these results suggest that down regulation of AP-1 complex gene expression may be driven in part by changing enhancer chromatin architecture.

AP-1 motif enrichment at differentially acetylated promoter-distal regions, along with a loss of AP-1 subunit gene expression, suggests that AP-1 binding at enhancers could be perturbed in uterine leiomyomas. To test this, ChIP-seq of AP-1 subunits JUN and FOS was performed (Table 3.1, Fig. 3.4.4a, b). Motif analysis of promoter-distal JUN and FOS bound sites confirmed that the majority of identified peaks overlap with AP-1 motifs (Fig. 3.4.5). 2,993 out of 25,736 FOS peaks were differentially bound in leiomyoma tissue samples with 86% of altered binding sites ($n = 2,573/2,993$) showing a loss in binding affinity (Fig. 3.4.6a, Fig. 3.4.7a). JUN binding was perturbed in the same manner, with 2,919 out of 18,283 JUN sites differentially bound and 95% of sites ($n = 2,773/2,993$) also showing a decrease in binding affinity (Fig. 3.4.6b, Fig. 3.4.7b). An investigation of all differentially bound FOS and JUN distal sites revealed an overlap with H3K27Ac signal, with a positive correlation between H3K27Ac signal changes and FOS/ JUN differential binding at these sites also observed (Fig. 3.4.8). Importantly, 761 enhancer regions with differentially bound FOS or JUN were identified as directly interacting with 420 differentially expressed gene promoters, which include extracellular matrix associated genes and a subset of previously identified AP-1 target genes with promoter-proximal AP-1 motifs (Fig. 3.4.9). AP-1 depleted enhancer sites that coincided with altered promoter contacts were also identified, with an increase in JUN and FOS binding affinity being associated with an increase in enhancer-promoter contact strength (Fig. 3.4.10a, b). In contrast, a decrease in binding affinity was associated with both a gain and a loss in enhancer-promoter
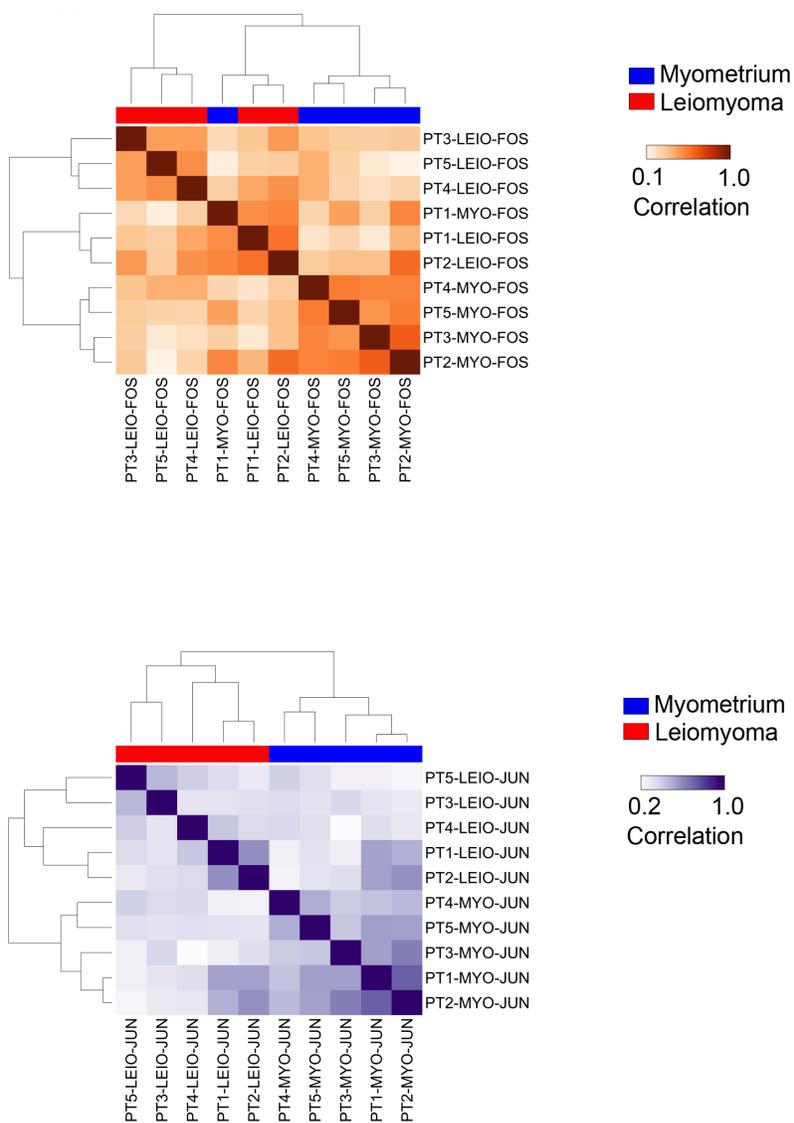
Fig. 3.4.4. Heat maps of AP-1 subunit ChIP sample affinity scores. (a, b) Correlation (Pearson) heat map of FOS (**a**) and JUN (**b**) ChIP sample affinity scores obtained from myometrium (blue) and leiomyoma (red) ChIP-seq read counts.

TF motif enrichment at promoter-distal
sites of JUN/FOS binding

Motif family	% sites	P value	Motif
Fra1	62.97%	1e-764	
Atf3	67.35%	1e-760	
Fra2	58.69%	1e-760	
JunB	61.86%	1e-747	
BATF	65.70%	1e-733	
Fosl2	48.91%	1e-722	
AP-1	67.08%	1e-691	

Fig. 3.4.5. AP-1 motif enrichment at AP-1 bound promoter-distal CREs. Top identified transcription factor motifs (*P* value ranking) enriched in enhancer regions bound by FOS or JUN in myometrium vs. leiomyoma.

a
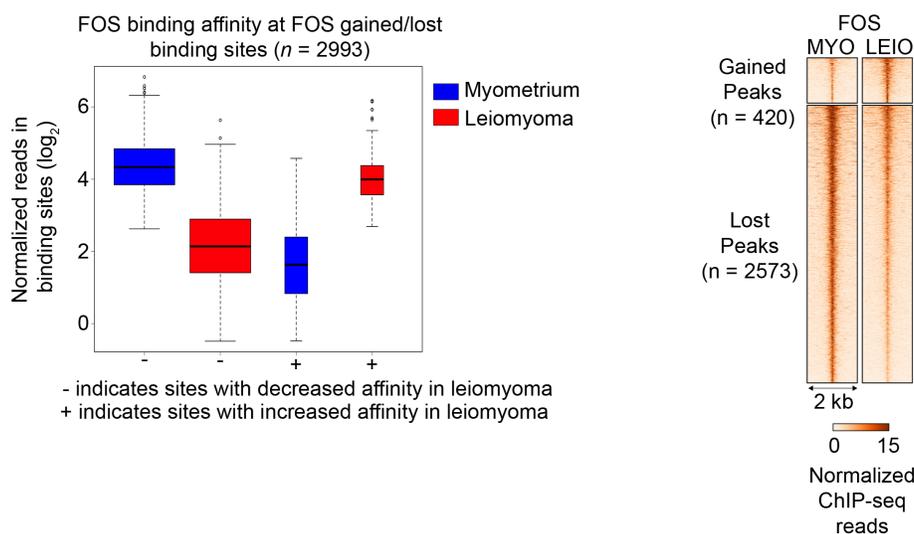
b
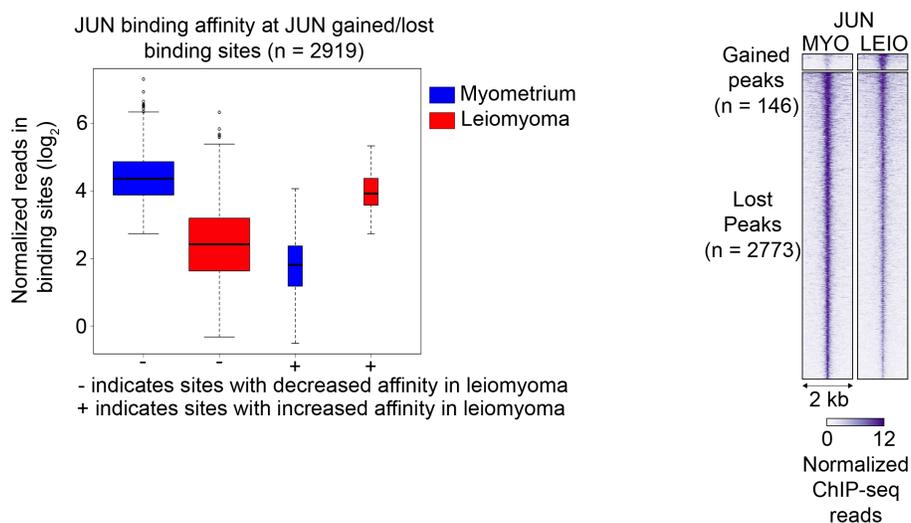
Fig. 3.4.6. Differential AP-1 occupancy in uterine leiomyomas. (a, b) Heat maps of normalised FOS (**a**) and JUN (**b**) ChIP-seq reads at differentially bound sites in myometrium vs. leiomyoma (right panels). Heat maps represent average signal of 5 biological replicates. Left panels show box and whisker plots of normalised reads at differentially bound FOS (**a**) and JUN (**b**) sites in myometrium and leiomyoma. Plots of sites with depleted (-) and enriched (+) FOS (**a**) and JUN (**b**) signal are shown separately. Whiskers are 1.5 x interquartile range (IQR).
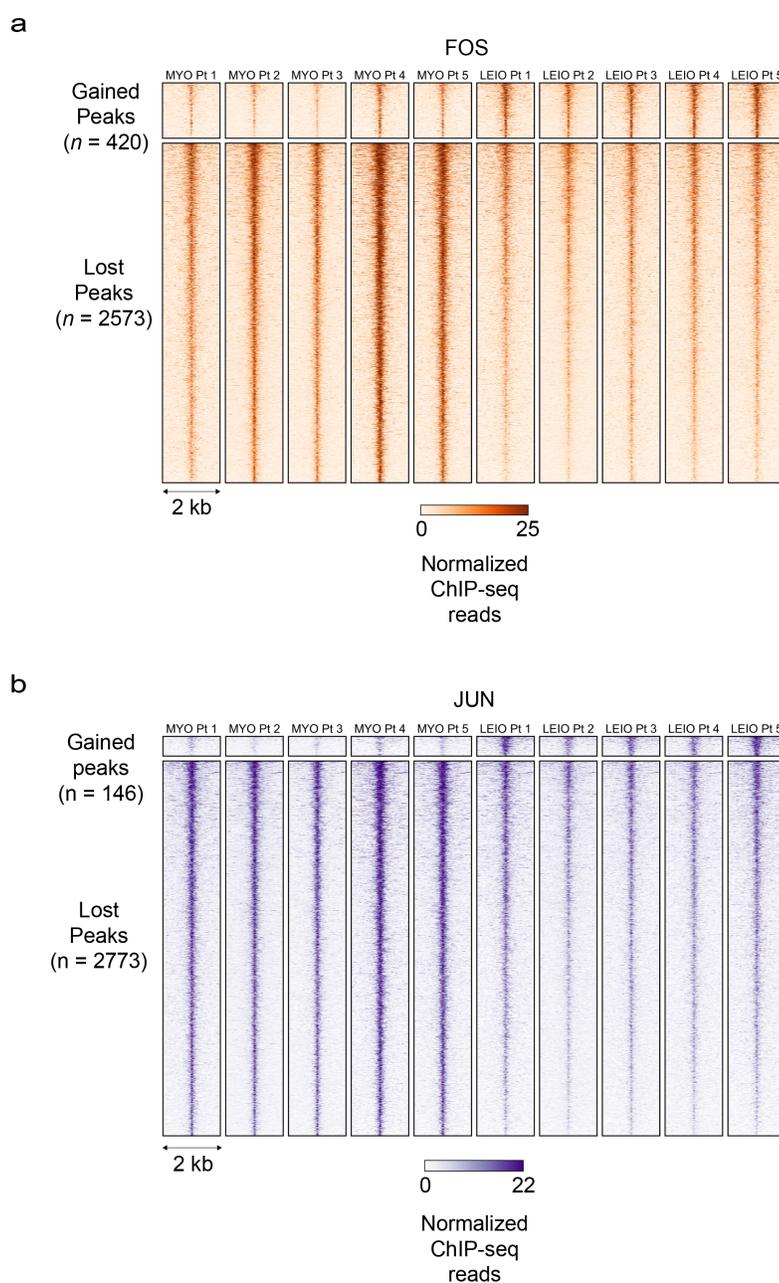
Fig. 3.4.7. AP-1 ChIP-seq reads at differentially bound regions in myometrium vs. leiomyoma. (a, b) Heat map of normalised FOS **(a)** and JUN **(b)** ChIP-seq reads at differentially bound FOS and JUN regions respectively in myometrium vs. leiomyoma. Signal for each biological replicate is shown.
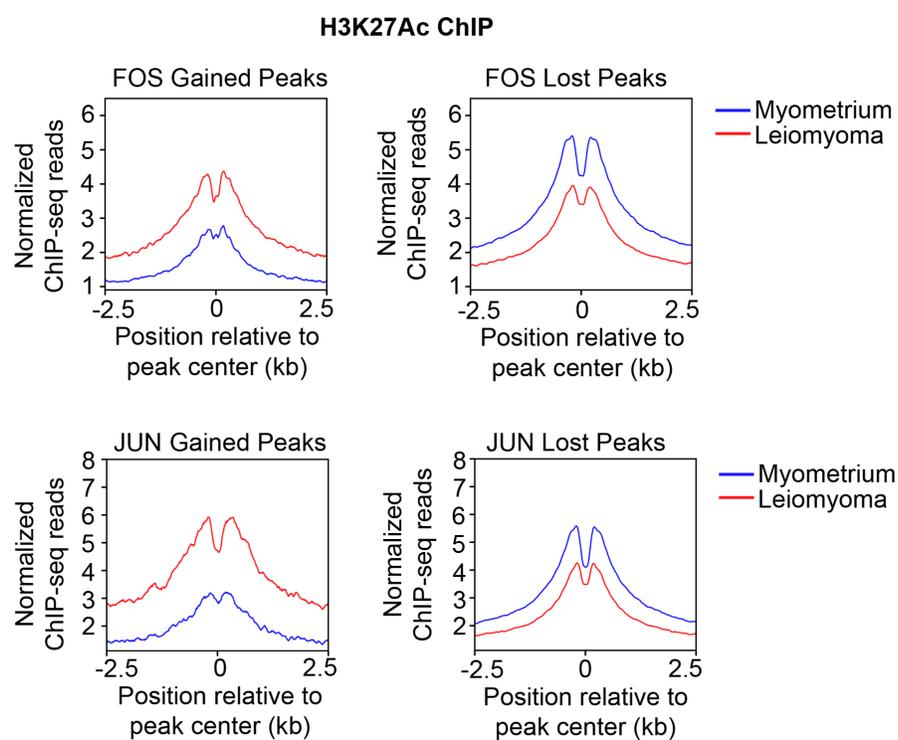
H3K27Ac ChIP

Fig. 3.4.8. Aggregate H3K27Ac signal at differentially bound AP-1 sites.
Tag density plots of H3K27Ac ChIP-seq enrichment profiles at FOS (top
panels) and JUN (bottom panels) gained and lost binding sites in myometrium
and leiomyoma.
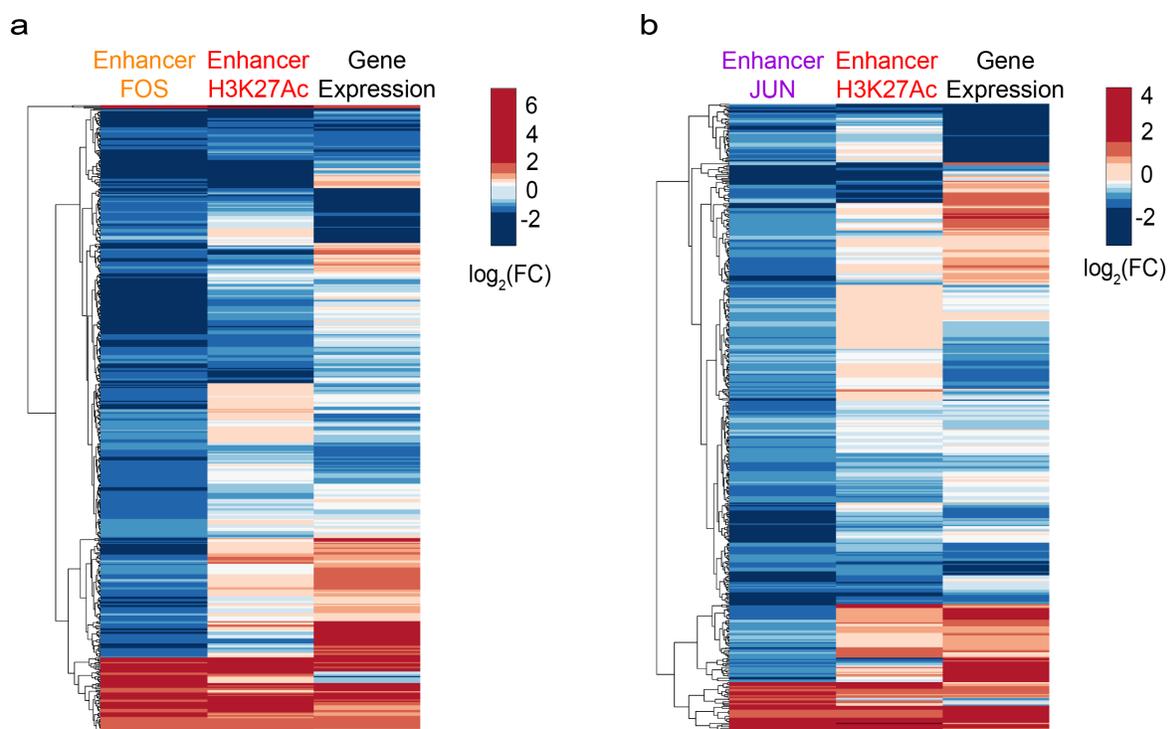
Fig. 3.4.9. Heat maps of changes in AP-1 occupancy at genomic loci associated with differentially expressed genes. Heat maps of changes in gene expression, enhancer H3K27Ac, and enhancer FOS (**a**) or JUN (**b**) binding associated with differentially expressed genes. Genes are hierarchically clustered by similarity in profiles of differential gene expression and FOS (n = 538) or JUN (n = 482) binding.
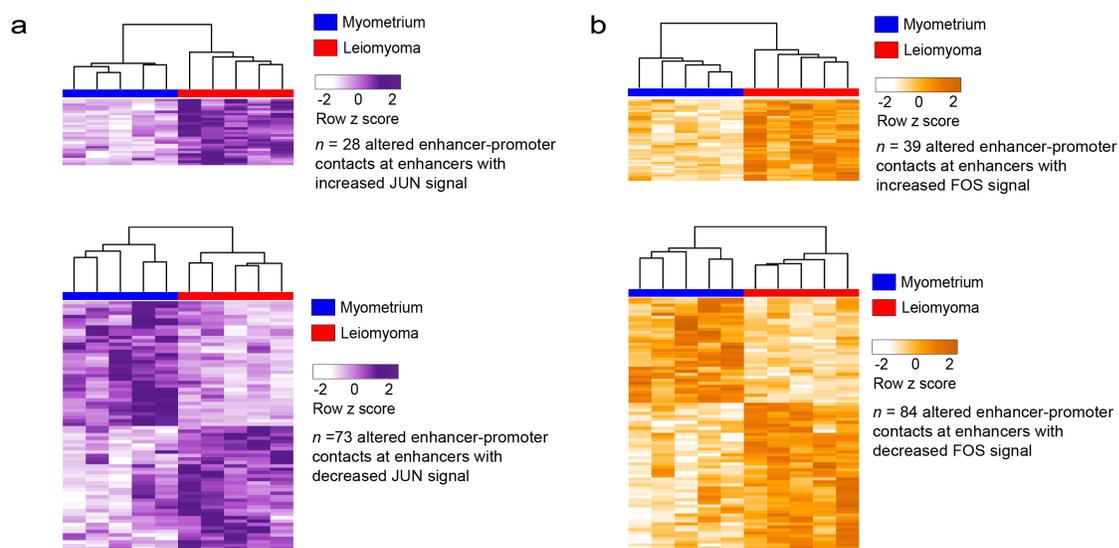
Fig. 3.4.10. Altered enhancer-promoter contacts at differentially bound AP-1 sites.
(**a**) Heat maps of altered enhancer-promoter contacts that overlap with enriched JUN (top panel, $n = 28$ contacts) and depleted JUN (bottom panel, $n = 73$ contacts) at enhancer regions in myometrium and leiomyoma tissues. (**b**) Heat maps of altered enhancer-promoter contacts that overlap with enriched FOS (top panel, $n = 39$ contacts) and depleted FOS (bottom panel, $n = 84$ contacts) at enhancer regions in myometrium and leiomyoma tissues.

contact strength. Examples of differentially expressed genes that show changes in enhancer binding of JUN and FOS include *ADAM19* and EGF-containing fibulin-like extracellular matrix protein 1 (*EFEMP1*) (Fig. 3.4.11a, b).

ChIP-seq of AP-1 subunits JUN and FOS in myometrium and leiomyoma tissue samples revealed a depletion of AP-1 on chromatin in leiomyomas, including depletion at a subset of enhancers that make contacts with promoters of dysregulated genes. From these results it is concluded that AP-1 occupancy at H3K27 acetylated enhancers in myometrium is important in regulating uterine muscle cell gene regulatory programs and a loss of AP-1 binding at enhancers as a result of decreased AP-1 gene expression may lead to widespread gene dysregulation in leiomyomas.

3.5. CDK8 subcomplex chromatin occupancy correlates with changes in enhancer acetylation

MED12 is a member of the four-subunit CDK8 submodule comprised of cofactors CDK8, MED12, MED13, and cyclin C. The CDK8 submodule is one of four Mediator modules and is important in the regulation of RNA polymerase II dependent gene transcription [404]. The submodule associates with the core Mediator complex to form a stable CDK8-Mediator complex which alters the transcription regulation function of Mediator [255]. Previously published interaction studies in heterologous cell lines expressing mutant MED12 (G44D/S) have suggested that mutations in exon 2 of MED12 may result in a decrease in CDK8 submodule binding affinity at sites of active transcription [273].

To assess whether exon 2 mutations at glycine 44 of MED12 result in a loss of CDK8 submodule chromatin binding in uterine leiomyomas, ChIP-seq of CDK8 and MED12 was
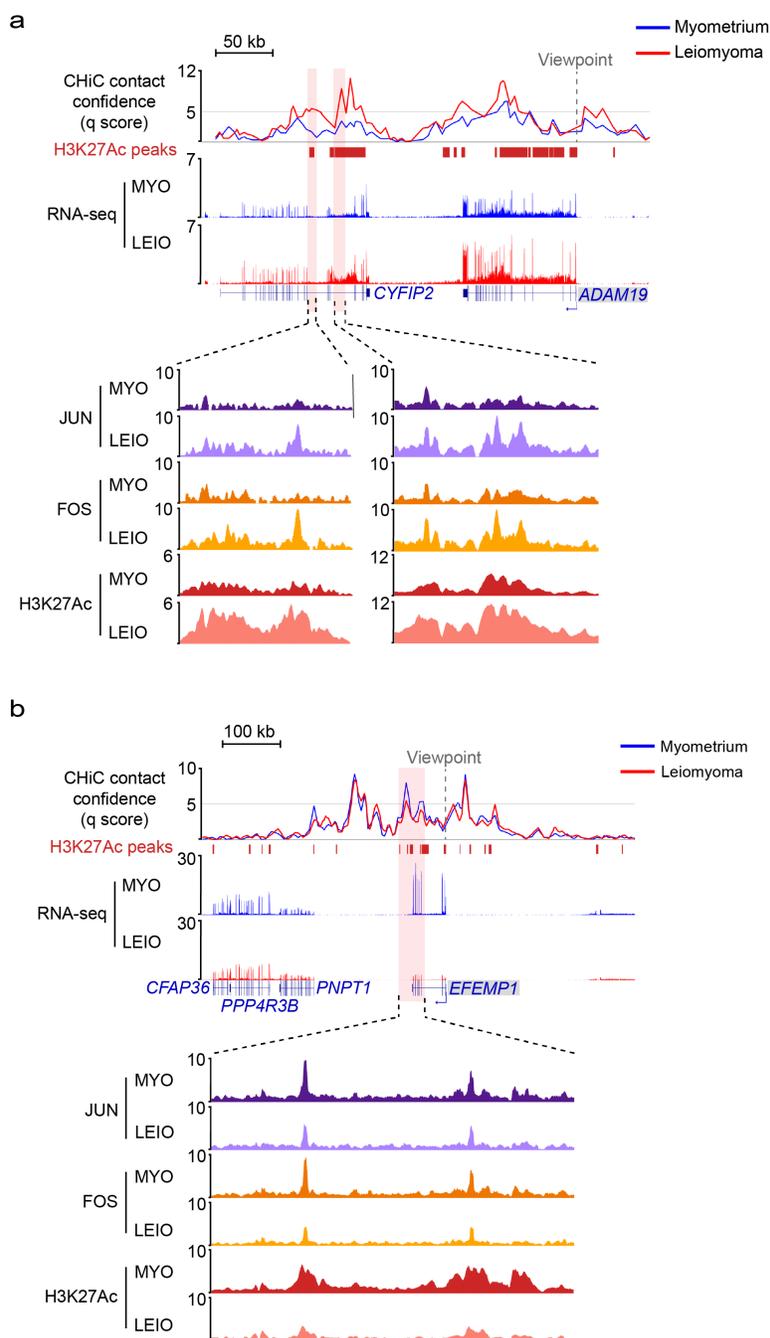
Fig. 3.4.11. Genomic loci containing differentially enriched AP-1 sites that are associated with dysregulated genes. (**a, b**) Differential enhancer AP-1 enrichment at *ADAM19* (**a**) and *EFEMP1* (**b**) genomic loci. Confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. In addition, RNA-seq genomic tracks are shown. ChIP-seq signal for H3K27Ac, FOS, and JUN in enhancer regions with differential FOS and JUN binding are highlighted (bottom zoomed in insert). Differential enhancer AP-1 binding occurs at enhancers with altered (**a**) and unaltered (**b**) enhancer-promoter contacts.

performed (Table 3.1, Fig. 3.5.1a, b). 30,697 CDK8 submodule binding sites co-bound by CDK8 and MED12 were identified in myometrium and leiomyoma tissue samples, with hierarchical clustering analysis revealing tissue type-specific CDK8 submodule occupancy in myometrium and leiomyoma (Fig. 3.5.2a, b). Despite a difference in occupancy profiles between myometrium and leiomyoma samples, a global loss of submodule chromatin binding affinity consistent with a MED12 mutation-driven binding defect was not observed. However, 18% ($n = 5,409/30,697$) of CDK8 submodule binding sites show changes in binding affinity in leiomyoma tissue samples as compared to myometrium (Fig. 3.5.3a, b, Fig. 3.5.4a, b). Unlike AP-1 loss of binding, submodule differential binding in leiomyoma samples was characterised by nearly equal numbers of increased and decreased binding sites, with 2,519 sites of submodule enrichment and 2,890 sites of submodule depletion. Also, changes in CDK8 submodule binding are positively correlated with changes in H3K27Ac signal at enhancers (Fig. 3.5.5a, b). This suggests that changes in CDK8 submodule chromatin occupancy reflect an altered enhancer landscape in diseased tissues rather than a global loss of CDK8 binding at sites of RNAPII mediated transcription as a result of MED12 exon 2 mutations.

Given the high correlation between H3K27 acetylation and CDK8 submodule binding at enhancers, the overlap between AP-1 and the CDK8 submodule at differential AP-1 enhancer binding sites was investigated. All JUN and FOS differentially bound sites were also co-bound by CDK8 and MED12 (Fig. 3.5.6, Fig. 3.5.7). Significantly, changes in CDK8 and MED12 binding affinities are positively correlated with changes in JUN and FOS binding at differentially bound JUN and FOS sites. AP-1 complex and CDK8 submodule co-occupancy is observed at altered enhancer regions of *FN1* and *ETS2* (Fig. 3.5.8, Fig. 3.5.9). Together, these results
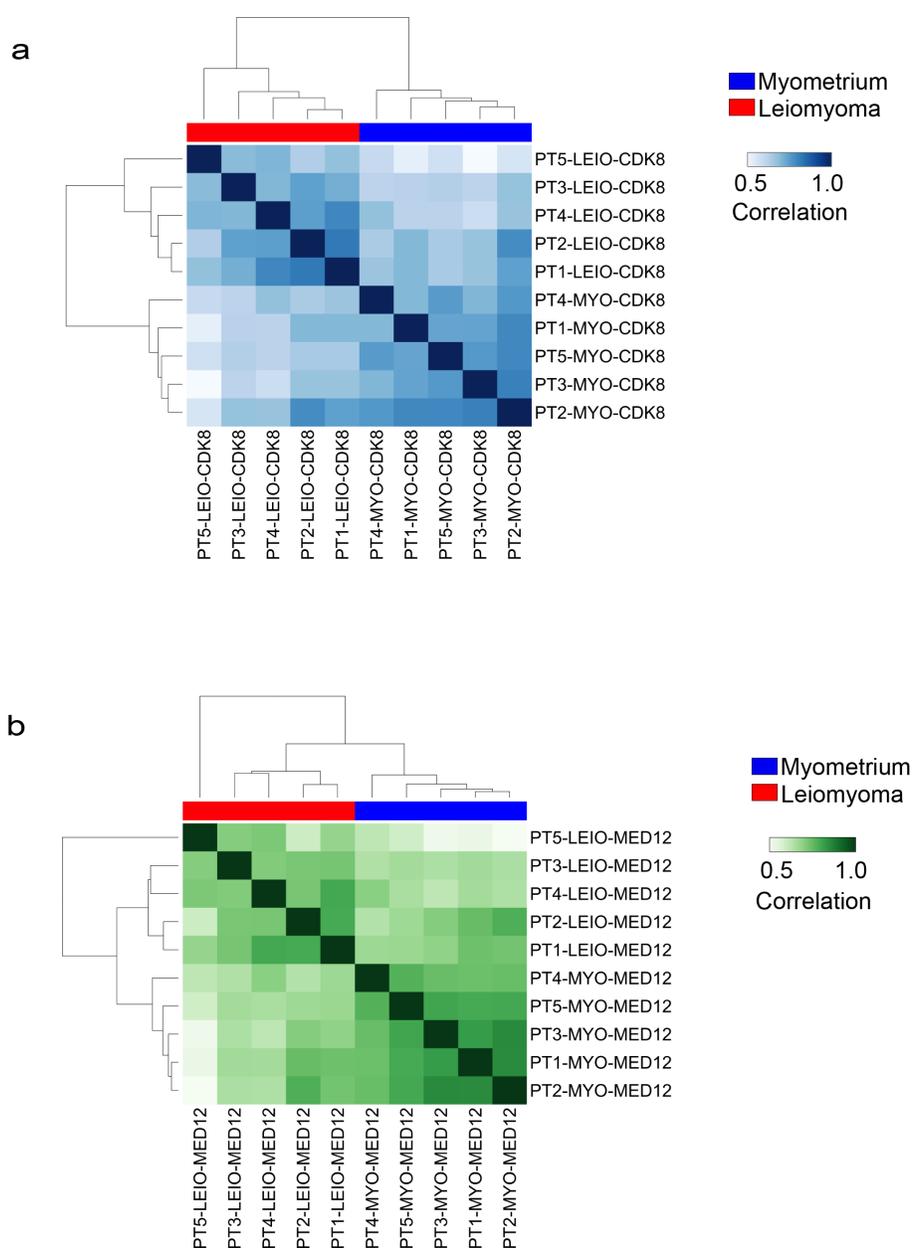
Fig. 3.5.1. Tissue type-specific CDK8 and MED12 ChIP occupancy. (a, b) Correlation (Pearson) heat map of CDK8 (**a**) and MED12 (**b**) ChIP sample affinity scores obtained from myometrium (blue) and leiomyoma (red) ChIP-seq read counts.
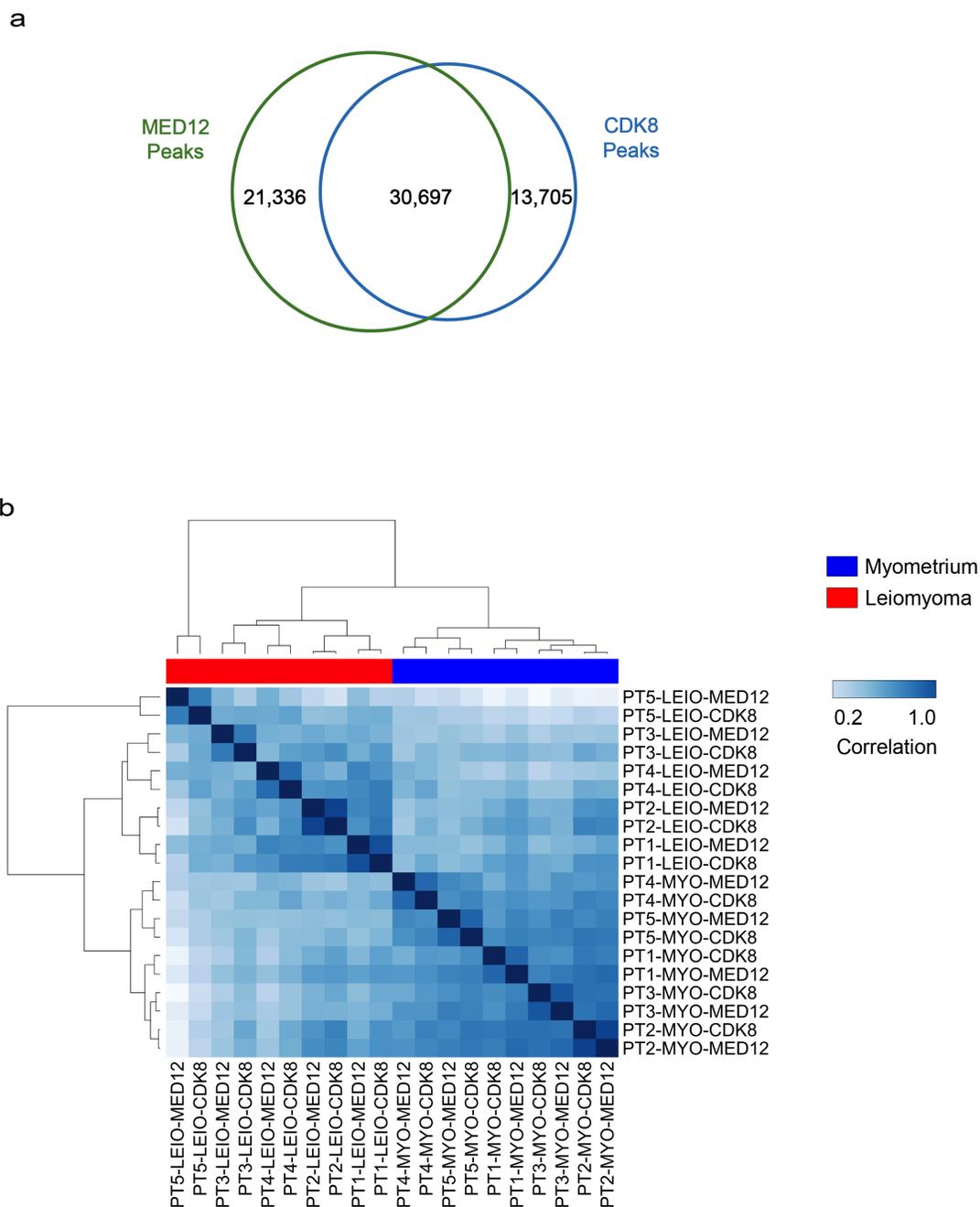
Fig. 3.5.2. CDK8 subcomplex chromatin occupancy profile in uterine leiomyomas.
(**a**) Venn diagram of overlapping CDK8 and MED12 peaks. Overlapping peaks were considered to be representative of CDK8 submodule peaks. (**b**) Correlation (Pearson) heat map of CDK8 and MED12 ChIP sample affinity scores at sites co-bound by CDK8 and MED12 in myometrium (blue) and leiomyoma (red).

156
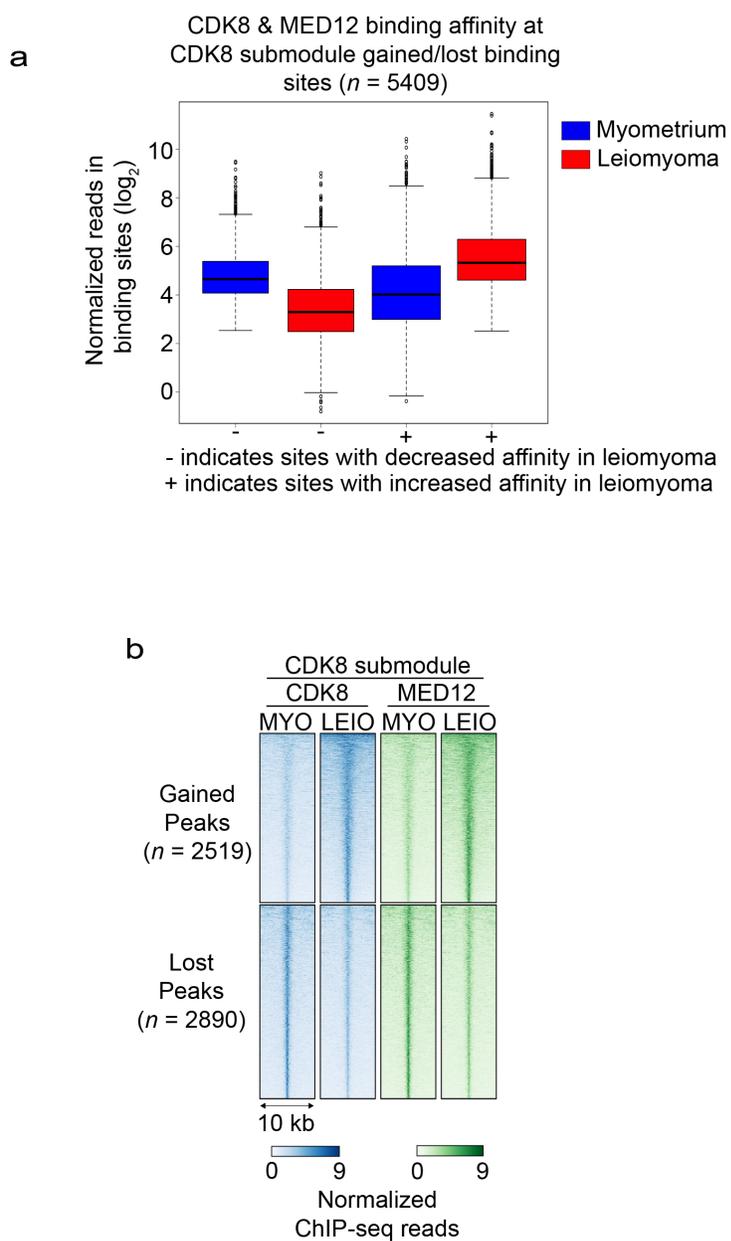
Fig. 3.5.3. Differential CDK8 submodule occupancy in uterine leiomyomas. (**a**) Box and whisker plot of normalised reads at CDK8 and MED12 co-bound sites that show enrichment or depletion of both factors. Whiskers are 1.5 x interquartile range (IQR). (**b**) Heat maps of normalised CDK8 and MED12 ChIP-seq reads at differentially bound sites in myometrium vs. leiomyoma. Heat maps represent average signal of 5 biological replicates.
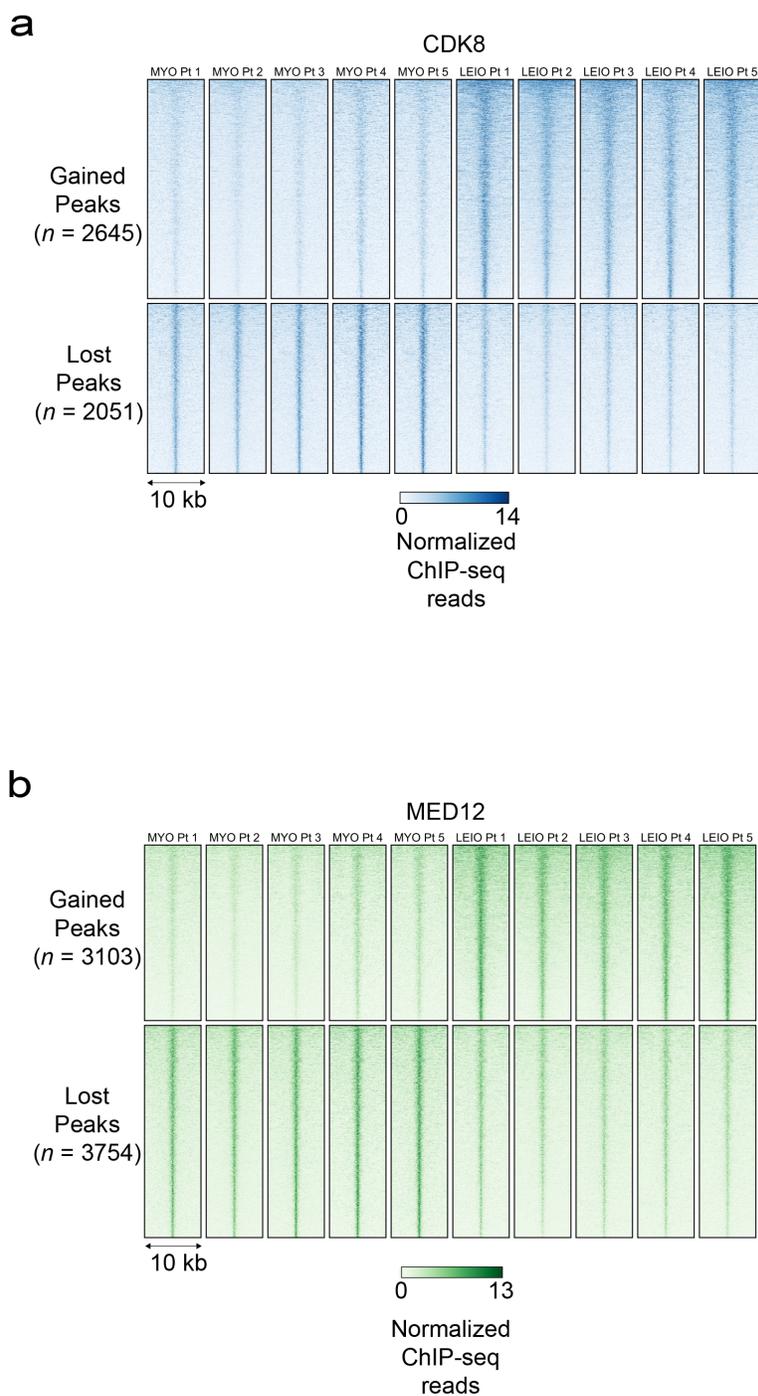
Fig. 3.5.4. CDK8 and MED12 ChIP-seq reads at differentially bound regions in myometrium vs. leiomyoma. (a, b) Heat map of normalised CDK8 (**a**) and MED12 (**b**) ChIP-seq reads at differentially bound CDK8 and MED12 regions respectively in myometrium vs. leiomyoma. Signal for each biological replicate is shown.
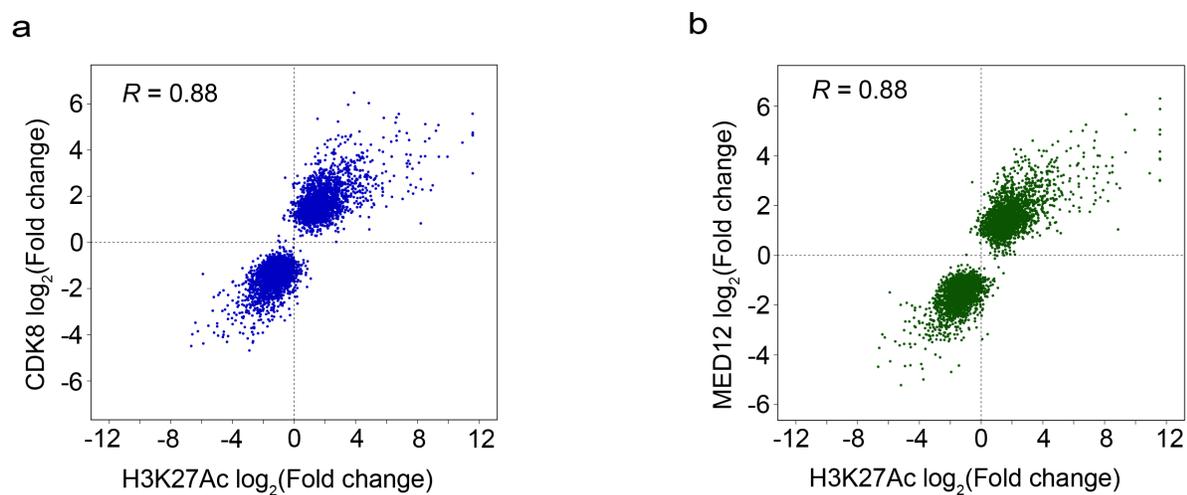
Fig. 3.5.5. Overlap between differential H3K27Ac signal and CDK8 submodule differential occupancy. (**a, b**) Scatter plots of fold change in H3K27Ac signal against fold change in CDK8 (**a**) and MED12 (**b**) at enhancers with enriched or depleted CDK8 and MED12 binding. *R* represents the Pearson correlation coefficient.

FOS Gained Peaks (*n* = 420)

FOS Lost Peaks (*n* = 2573)
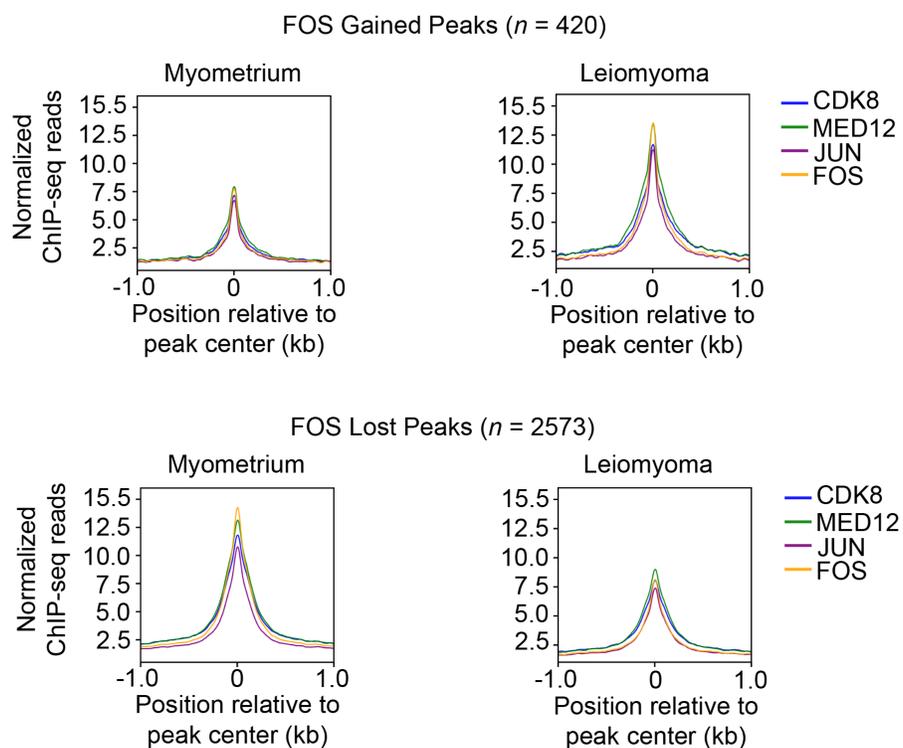
Fig. 3.5.6. AP-1 and CDK8 submodule co-occupancy at regions of differential FOS occupancy. Tag density plots of CDK8, MED12, JUN, and FOS ChIP-seq enrichment profiles at FOS enhancer binding sites in myometrium and leiomyoma.
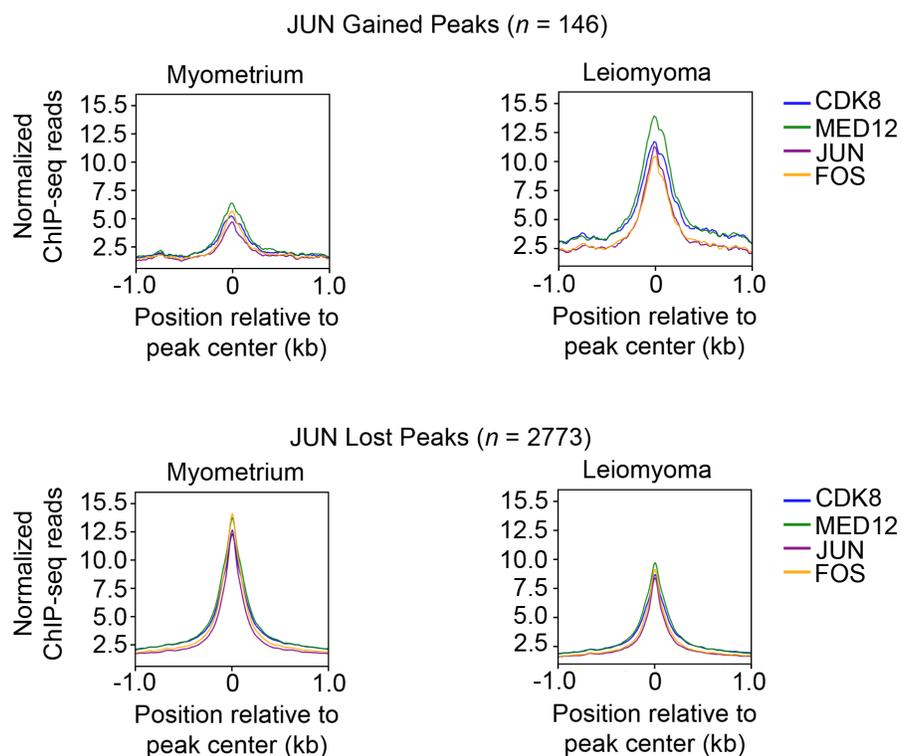
Fig. 3.5.7. AP-1 and CDK8 submodule co-occupancy at regions of differential JUN occupancy. Tag density plots of CDK8, MED12, JUN, and FOS ChIP-seq enrichment profiles at JUN enhancer binding sites in myometrium and leiomyoma.
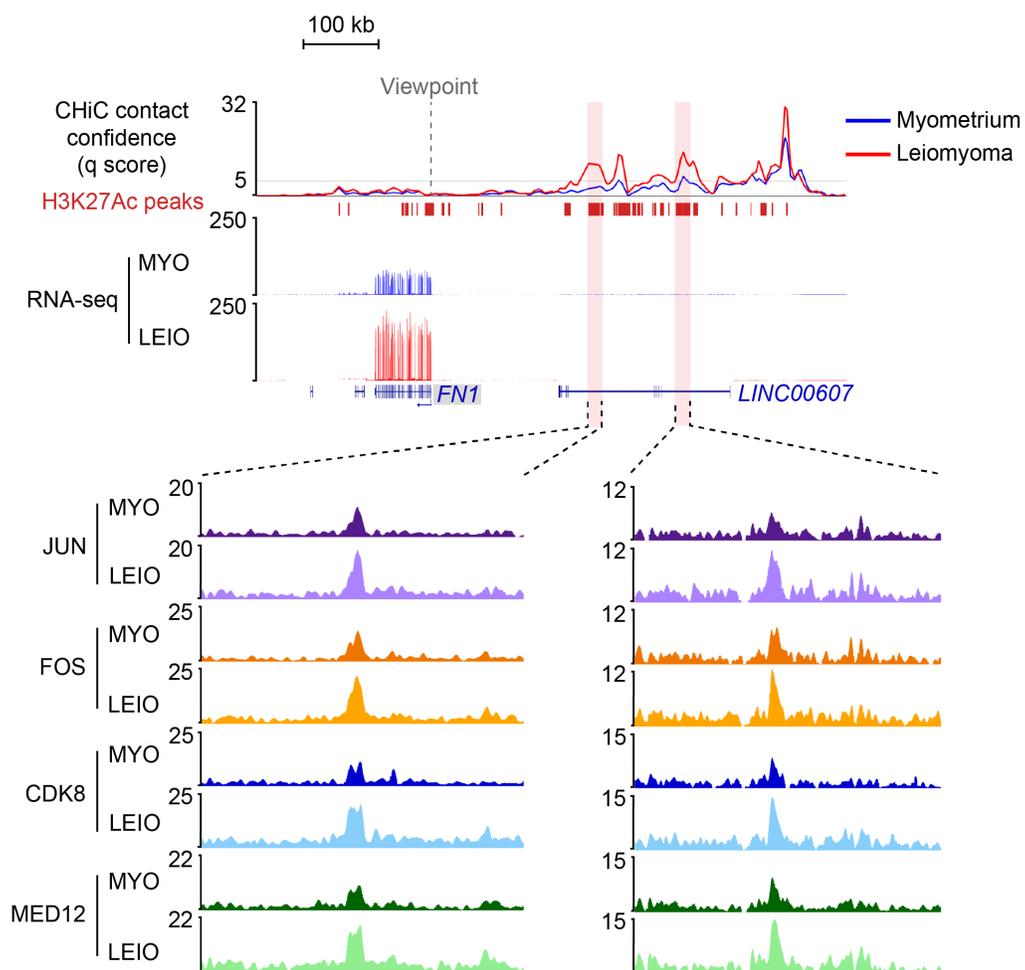
Fig. 3.5.8. AP-1 and CDK8 submodule co-occupancy at genomic loci associated with dysregulated genes. AP-1 subunits (JUN and FOS) and CDK8 submodule (CDK8 and MED12) binding profiles at *FN1* associated genomic loci. Confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. Enhancer-promoter contacts and RNA-seq genomic tracks for myometrium and leiomyoma are also shown.
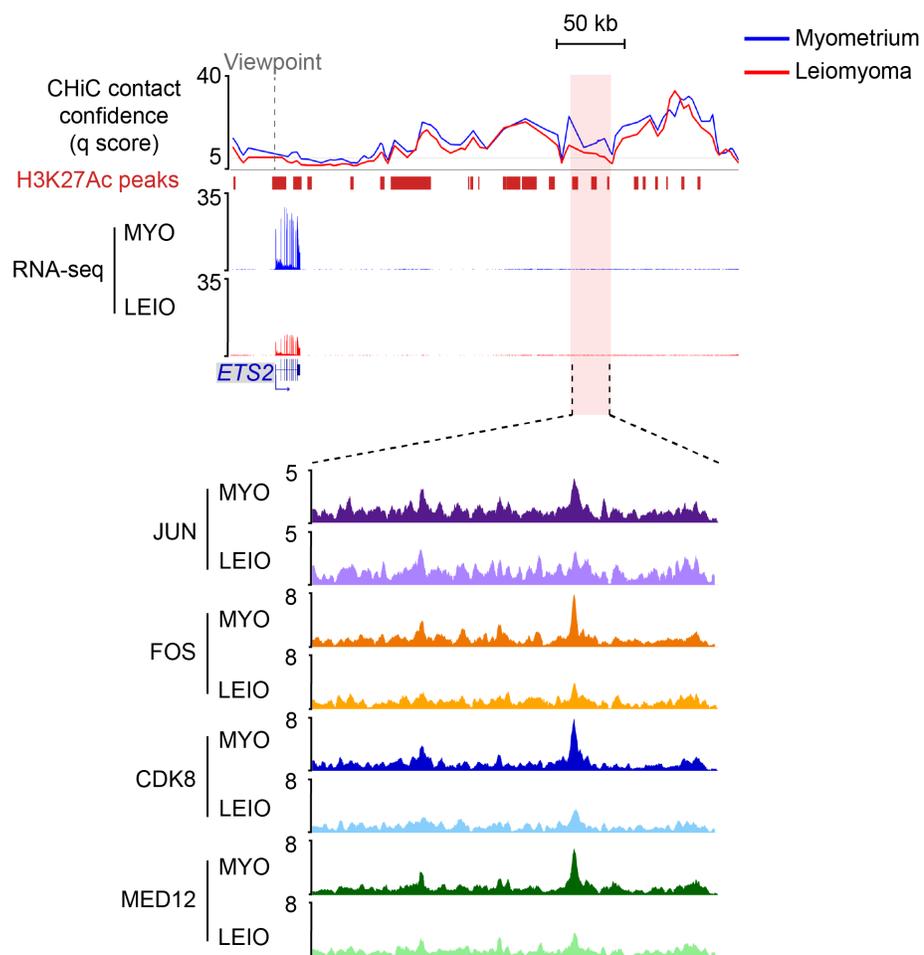
Fig. 3.5.9. AP-1 and CDK8 submodule co-occupancy at genomic loci associated with dysregulated genes. AP-1 subunits (JUN and FOS) and CDK8 submodule (CDK8 and MED12) binding profiles at *ETS2* associated genomic loci. Confidence scores (CHiCAGO *q* scores) are shown for myometrium and leiomyoma. Enhancer-promoter contacts and RNA-seq genomic tracks for myometrium and leiomyoma are also shown.

establish that global loss of CDK8 submodule chromatin binding is not an epigenetic feature of MED12 mutant leiomyomas. Rather, alterations in enhancer chromatin state, characterised by changes in H3K27Ac signal, promoter contact strength, and altered AP-1 complex/ CDK8 submodule chromatin binding, together may account for dysregulated gene expression in uterine leiomyomas.

3.6. Loss of AP-1 in human uterine smooth muscle cells leads to altered enhancer architecture and extracellular matrix gene dysregulation

The identification of direct AP-1 target genes among dysregulated genes in leiomyoma tissue samples, in concert with the down regulation in AP-1 subunit gene expression and the depletion of JUN/FOS on chromatin, together suggest that AP-1 may be an important protein complex in myometrium gene regulation whose aberrant expression may play a role in leiomyoma disease pathogenesis. To test this, *JUN* and *FOS* family of genes were silenced in primary human uterine smooth muscle cells (HUtSMC) by CRISPR/Cas9 mediated protein depletion (Fig. 3.6.1). RNA sequencing of AP-1 depleted biological replicate samples showed reproducible results and segregation dependent on AP-1 depletion or control (Fig. 3.6.2). Importantly, 1,894 genes were differentially expressed in AP-1 depleted cells, with gene ontology analysis revealing changes in expression of genes involved in extracellular matrix organisation (Fig. 3.6.3 and Fig. 3.6.4). These results were further complemented by AP-1 subunit gene silencing using lentiviral short hairpin RNAs followed by RNA sequencing in myometrium primary cells obtained from fresh patient tissue samples (Fig. 3.6.5a). This approach also resulted in dysregulation of genes shown
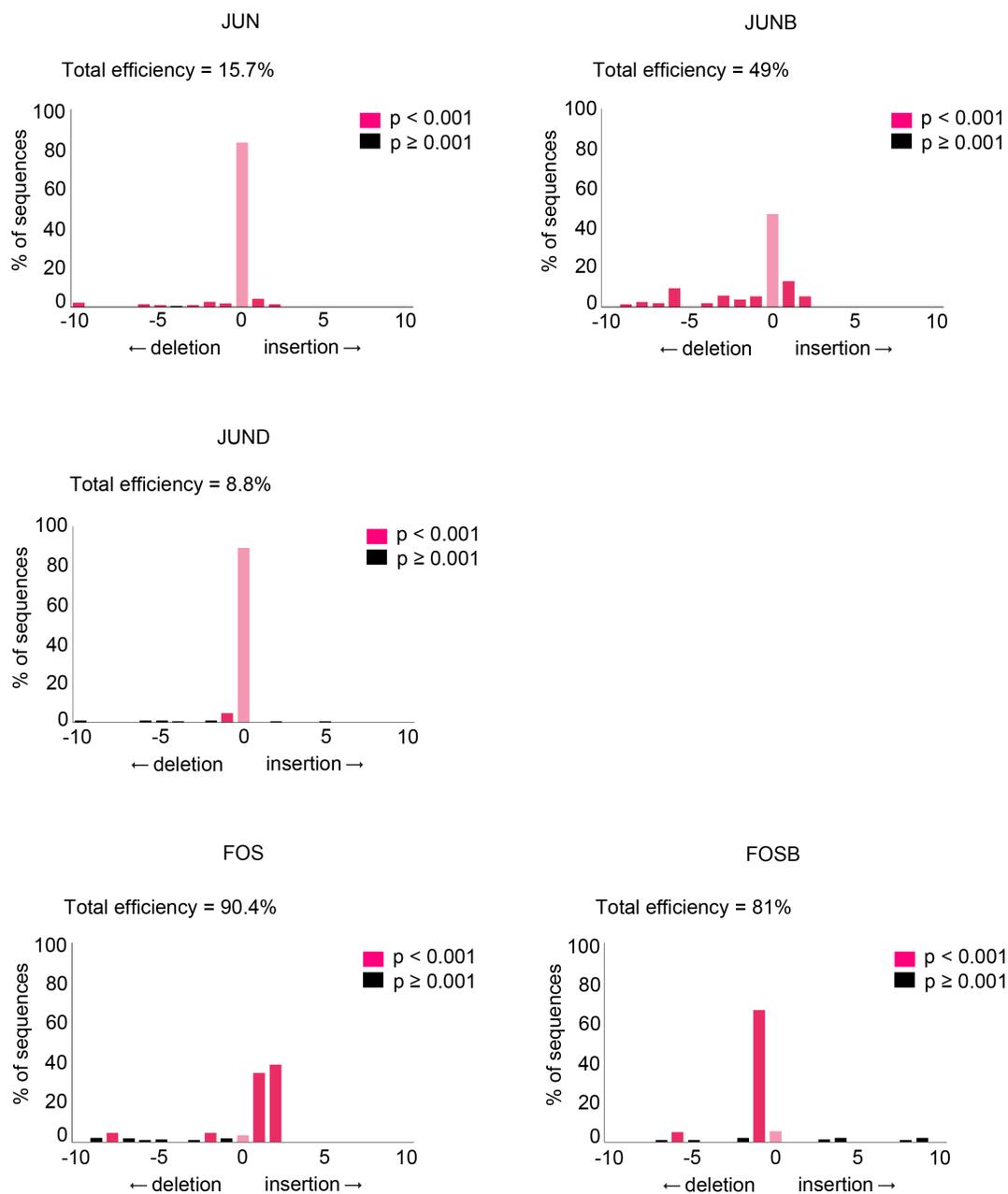
Fig. 3.6.1. CRISPR/Cas9 mediated gene silencing of AP-1 subunits *JUN*, *JUNB*, *JUND*, *FOS*, and *FOSB*. CRISPR/Cas9 mediated AP-1 subunit gene editing efficiencies as determined by sequence trace decomposition of sanger sequenced PCR amplicons.
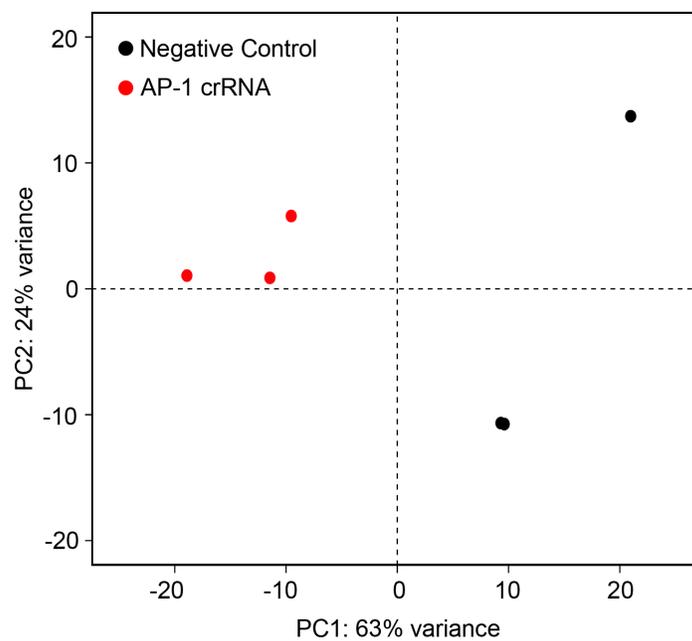
Figure 3.6.2. AP-1 loss accounts for the majority of the variance between AP-1 depleted and control cells. Principal component analysis plot of the top 500 most variable genes in CRISPR/Cas9 negative control and AP-1 depleted HUtSMCs. First two principal components are plotted.
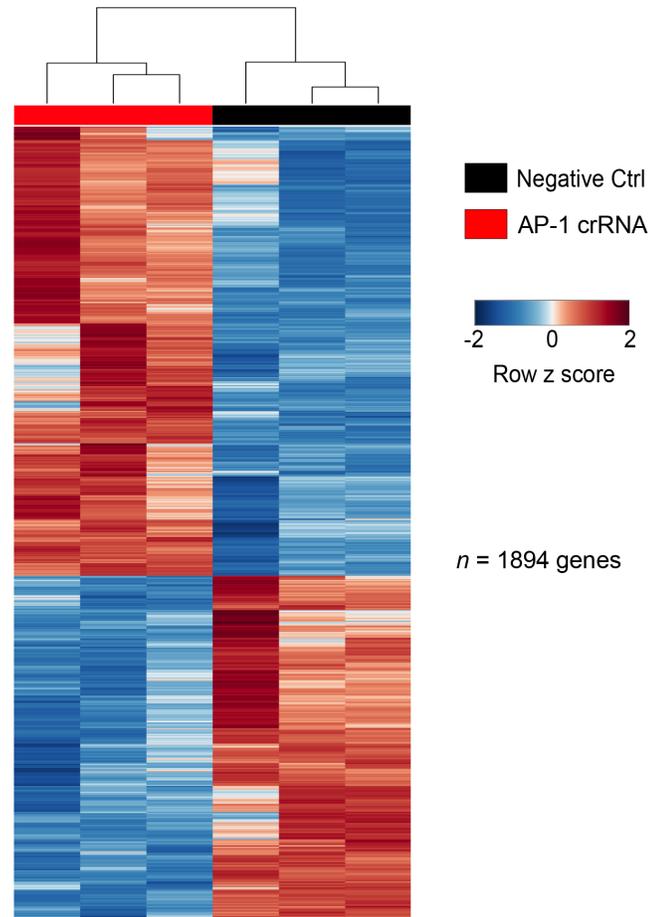
Figure 3.6.3. Heat map of differentially expressed genes in AP-1 depleted HUtSMCs. Gene expression levels in negative control (black) vs. AP-1 depleted (red) cells relative to the mean expression are shown as row z scores (n = 1,894 genes, FDR < 0.05).
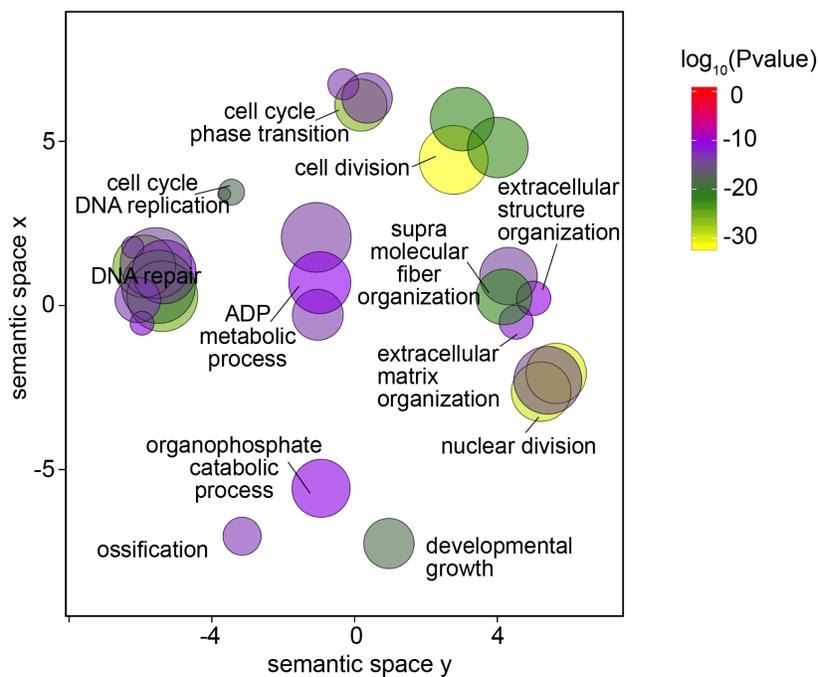
Figure 3.6.4. Differentially expressed genes in AP-1 depleted cells are involved in extracellular matrix formation, organisation and degradation. Scatter plot of confidence scores for enriched gene ontologies associated with differentially expressed genes in AP-1 depleted cells, with ontologies clustered by functional similarity in the semantic space.
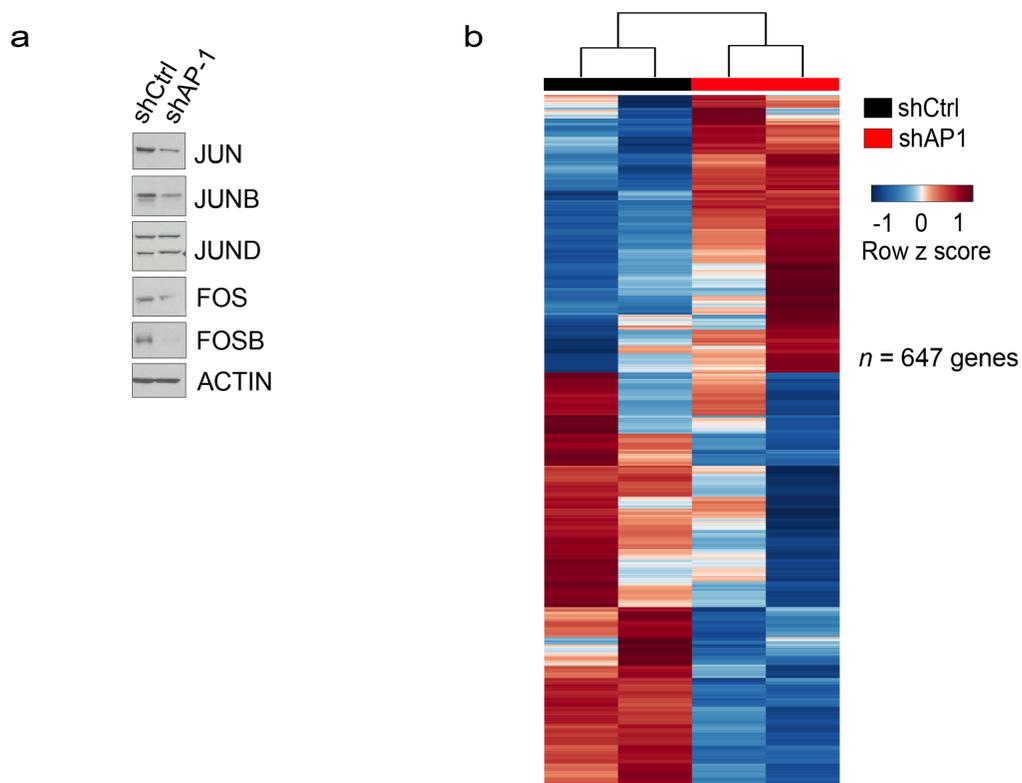
Fig. 3.6.5. Gene dysregulation in patient-derived primary cells upon AP-1 knockdown. (**a**) Western blot of AP-1 subunits in myometrium primary cells transduced with lentiviral hairpins targeting *JUN*, *JUNB*, *JUND*, *FOS* and *FOSB* (shAP-1). Western blot of cells transduced with non-targeting lentiviral hairpin control (shCtrl) are also shown. (**b**) Hierarchically clustered heat map of all differentially expressed genes in AP-1 knockdown primary cells. Gene expression levels relative to the mean expression are shown as row z scores.

to be involved in key biological processes that are also perturbed in uterine leiomyomas (Fig. 3.6.5b, Fig. 3.6.6). These results from parallel studies demonstrate that AP-1 loss in uterine muscles cells leads to large-scale changes in gene expression as also seen in leiomyomas.

The enrichment of AP-1 motifs at distal sites that show differential H3K27 acetylation as well as the overlap between differentially acetylated regions and sites of altered AP-1 occupancy observed in leiomyoma tissue samples suggests that AP-1 may play a role in enhancer malfunction in leiomyomas by directing acetylation states at distal sites. To test this, H3K27Ac ChIP sequencing was performed on AP-1 depleted HUtSM cells. 3,727 differentially acetylated regions were identified in AP-1 depleted cells as compared to negative controls (Fig. 3.6.7). Changes in H3K27 acetylation patterns of AP-1 depleted cells relative to negative controls were strikingly similar to acetylation changes observed in leiomyoma tissues as compared to myometrium. Similarly to differential H3K27 acetylation patterns observed in leiomyoma tissue samples, roughly equal proportions of sites showed a gain or loss in H3K27Ac signal. In addition, the majority of differentially acetylated regions were at promoter-distal sites, with a higher than expected number of differentially acetylated intergenic regions (Fig. 3.6.8, $\chi^2(4) = 346.7$, $p < 0.05$). Motif enrichment analysis also revealed almost exclusively AP-1 motifs as the top enriched motifs at differential H3K27Ac regions, with AP-1 motifs enriched at H3K27Ac peak centres (Fig. 3.6.9a, b).

Large-scale gene expression and promoter-distal H3K27Ac changes in AP-1 depleted cells demonstrate the importance of AP-1 in regulating enhancer acetylation and their associated genes. The similarity in profiles between AP-1 depleted cells and leiomyoma tissues samples strongly suggests that a significant part of the transcriptomic and epigenomic changes seen in

uterine fibroids may be directly linked to the loss in AP-1 gene expression. Taken together, these results suggest an important role played by AP-1 in uterine muscle cell enhancer regulation, the perturbation of which results in significant changes in enhancer regulation and gene expression, including the dysregulation of extracellular matrix associated genes, a hallmark of uterine leiomyoma disease pathogenesis.
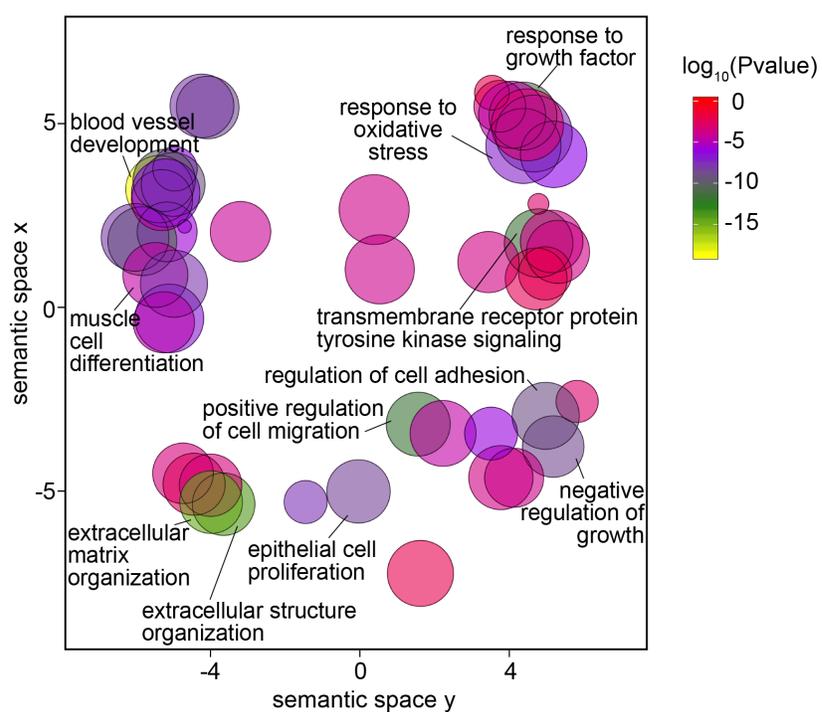
Fig. 3.6.6. Differentially expressed genes in AP-1 knockdown primary cells are involved in extracellular matrix formation, organisation and degradation. Scatter plot of confidence scores for enriched gene ontologies associated with differentially expressed genes in AP-1 knockdown primary cells, with ontologies clustered by functional similarity in the semantic space.
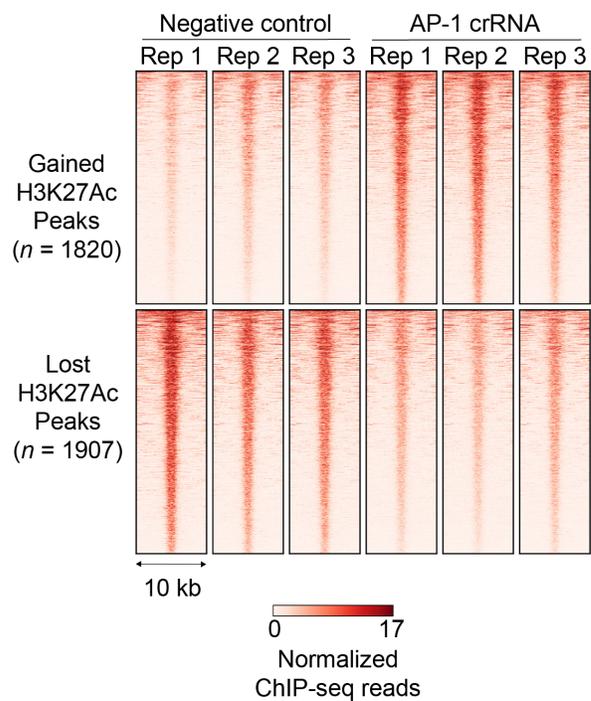
Figure 3.6.7. Heat map of H3K27Ac ChIP-seq reads at differentially acetylated regions in AP-1 depleted HUtSMCs. Normalised signal for each biological replicate is shown.

Negative control

Promoter-TSS (17.6%)
Exon (4.4%)
Intron (50.6%)
TTS (4.4%)
Intergenic (23%)

Total = 50678

AP-1 crRNA

Promoter-TSS (17.8%)
Exon (4.3%)
Intron (50.1%)
TTS (4.3%)
Intergenic (23.5%)

Total = 50853

Differential enrichment

Promoter-TSS (8.2%)
Exon (3.1%)
Intron (52.3%)
TTS (3.7%)
Intergenic (32.8%)

Total = 3727
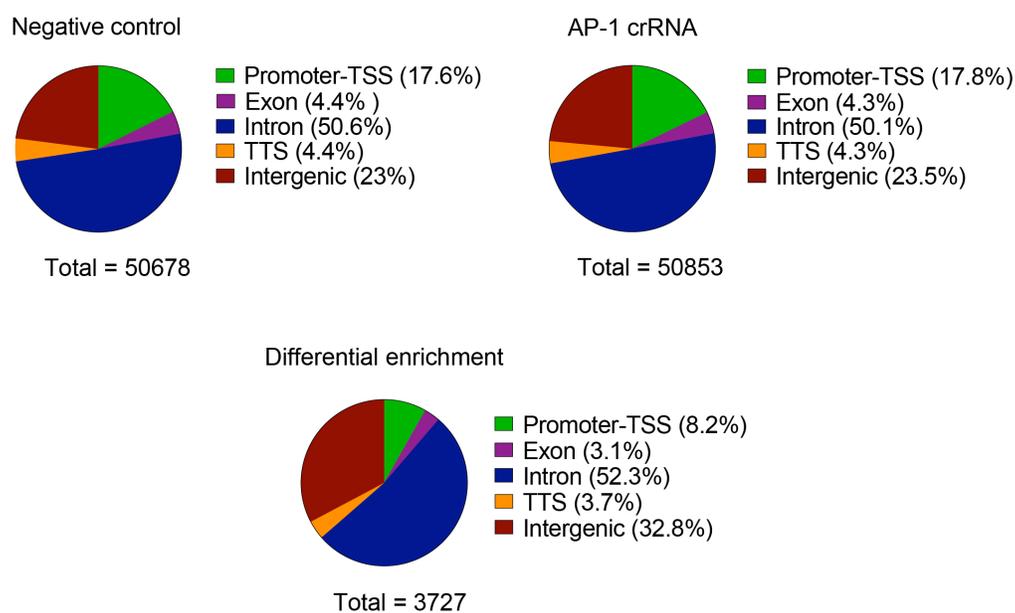
Fig. 3.6.8. Histone acetylation changes occur disproportionately at intergenic regions in AP-1 depleted HUtSMCs. (a) Pie charts of H3K27Ac signal distribution at genomic loci in HUtSMC negative control (top left panel), AP-1 depleted (top right panel) and at differentially acetylated regions (bottom centre panel).

a

TF motif enrichment at sites with differential H3K27 acetylation

Motif family	% sites	P value	Motif
Fra2	57.02%	1e-143	
FOS::JUNB	53.07%	1e-142	
Fosl2	42.63%	1e-137	
Fra1	63.40%	1e-129	
JUNB	61.26%	1e-121	

b
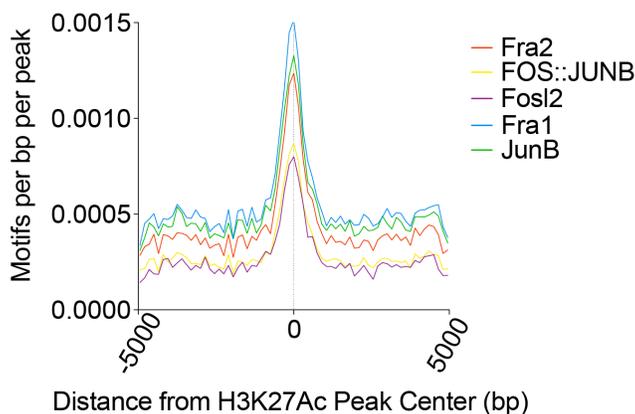
Fig. 3.6.9. AP-1 motif enrichment at differentially acetylated enhancer regions in AP-1 depleted HUtSMCs. (a) Top identified transcription factor motifs (*P* value ranking) enriched in enhancer regions with differential H3K27Ac signal between AP-1 depleted and negative control cells. (**b**) Motif enrichment plot at differential H3K27Ac peak centres in AP-1 depleted vs. negative control cells.

CHAPTER 4. DISCUSSION

Epigenomic profiling of blood disorders directly from patient samples has provided extensive insights into the pathogenesis of these diseases [405]. This has been aided primarily by the ease of performing genome-wide analyses directly from blood samples. However, epigenomic characterisation of solid tumours has remained challenging, necessitating the digestion of tumours and culturing of primary cells before profiling. Recent studies have demonstrated the shortcomings of this strategy with regard to uterine leiomyomas [367]. In culture, a rapid decrease of the *MED12* mutant leiomyoma cell population is observed, with a near complete loss of *MED12* mutant cells seen as early as the third passage after initial *in-vitro* culture of cells obtained from collagenase-digested tissue samples. A reduction in the *MED12* mutant leiomyoma cell population is also observed without passaging of cultured cells, with a decrease in mutant cells seen in cultures allowed to grow to and past complete confluence, suggesting that *in vitro* conditions are generally unfavourable for leiomyoma cell growth [368].

In vitro spheroid culture of leiomyoma cells obtained from digested patient tissue samples has been proposed as a viable alternative to conventional cell culture techniques [406-408]. However, the survivability of leiomyoma cells in this model has not been ascertained and it is unclear if this model performs better than conventional cell culture with regard to leiomyoma cell viability. For example, although Xie et al. performed spheroid culture of *MED12* mutant leiomyoma cells, the mutational status of the cell population throughout the experiment was not monitored. Hence it is unknown if *MED12* mutant cells continue to proliferate or, similar to conventional *in vitro* cell culture, this population of cells decreases over time [367,368,406]. In general, primary cell culture studies highlight the unsuitability of current leiomyoma cell models

in the study of uterine leiomyoma disease pathogenesis and necessitate the use of alternative models in the study of this disease.

Using modified methods adapted for use with fresh tissue samples without the need for isolation and culturing of primary cells, this study provides an extensive transcriptomic and epigenomic characterisation of normal myometrium and *MED12* mutant leiomyoma tissues, thereby yielding a near-faithful snapshot of the differences between myometrium and leiomyoma tissues under native conditions. These methods should provide a reproducible approach for the characterisation of solid tumours in general without the need for primary cell culture.

Through comprehensive profiling of transcribed genes in myometrium and leiomyoma tissue, two distinct gene expression profiles for myometrium and leiomyoma samples were identified. A high degree of similarity among biological replicates implies a consistent mechanism of disease pathogenesis in *MED12* mutant uterine fibroids, which may be driven by the mutation. Nearly six thousand differentially expressed genes were identified in leiomyoma samples with protein-coding genes accounting for the majority of the altered transcriptome. The predominant feature revealed by gene ontology analysis is a perturbation of genes that are known to be important in the regulation of the extracellular matrix. Approximately a third of all known matrisome associated genes were found to be dysregulated in leiomyoma tissue samples and this significant enrichment for matrisome genes is consistent with leiomyoma disease pathology. Interestingly, although many genes coding for proteins involved in ECM degradation were enriched in the altered leiomyoma tissue transcriptome, many of these demonstrated an increase in expression as opposed to the expected decrease, ruling out a down regulation of matrix degrading processes as a mechanism of increased ECM deposition in leiomyomas.

It is important to note that although a third of all known ECM-associated genes are dysregulated in leiomyomas, this only accounts for approximately 5.5% of all differentially expressed genes identified in the leiomyoma transcriptome. RNA-seq analysis also identifies other possible biological processes that are perturbed in leiomyomas, which include the regulation of hormone levels and the regulation of cell secretion. Given the very significant proportion of the transcriptome that is altered in uterine leiomyomas, it will be key going forward to characterise the role of the many dysregulated genes and corresponding biological processes in uterine leiomyoma pathobiology.

The role of TGF-β signalling in uterine leiomyoma pathogenesis has previously been described [132,145,179,180]. In particular, research has focused on the role of the TGF-β subfamily in the disease, with TGF-β3 being identified as a possible target of GnRH agonists [181-185]. However, this study identifies the bone morphogenetic protein subfamily and the closely related growth and differentiation factor subfamily as highly dysregulated in leiomyomas. This suggests that BMP/GDF signalling may play an important role in ECM dysregulation in leiomyomas and future work to characterise the mechanism by which BMP/GDF signalling is perturbed in leiomyomas will be important.

Chromatin immunoprecipitation of acetylated histone H3K27 (H3K27Ac), a mark of active transcription, followed by massively parallel sequencing revealed substantial changes in H3K27 acetylation profile in leiomyoma tissue as compared to myometrium, with approximately 30% of all identified H3K27 acetylated regions showing changes in H3K27Ac signal. Notably, the largest alterations were seen at promoter-distal sites, with H3K27Ac signal changes at intergenic regions being the most prominent. Also, a significant number of promoters of

differentially expressed genes showed little to no change in H3K27 acetylation. Given these changes at predominantly promoter-distal CREs and the relative stability of H3K27Ac signal at some promoters, the data suggests a role for modifications in enhancer architecture as a mechanism for gene dysregulation in uterine leiomyomas. Overall, the very large changes in H3K27Ac cistrome between myometrium and leiomyoma tissues are commensurate with the equally large changes in tissue transcriptomes.

In addition to histone modifications, recently described promoter capture Hi-C allows for the assessment of intradomain chromatin architecture changes in cells, which include signal-dependent changes and changes driven by single nucleotide polymorphisms in distal *cis*-regulatory elements [364,365]. Promoter capture Hi-C carried out on patient tissue samples in conjunction with ChIP-sequencing of H3K27Ac revealed novel enhancer regulatory mechanisms in uterine fibroids. In addition to modified H3K27Ac signal enrichment at enhancers, a subset of contacts were altered in uterine leiomyomas. This observation of altered contacts in disease tissue samples is a novel finding and although the significance of altered contacts in the disease state is unclear, dynamic contacts have been demonstrated to be important chromatin architecture re-wiring events in cell differentiation and may therefore be a significant re-wiring event in leiomyomas [364,365].

A survey of enhancers associated with differential gene expression identified three classes of remodelled enhancers on the basis of changing H3K27 acetylation and altered promoter contact strength, of which enhancer regions with stable contacts and differential acetylation were the predominant class of altered enhancers. This is consistent with previous studies showing the lineage-defining characteristics of most promoter contacts, hence their

relative stability [402]. In addition to a significant overlap between genomic regions forming contacts with promoters and regions with H3K27 acetylation, some enhancer-promoter contacts that do not overlap H3K27 acetylated regions were also identified. Previous studies have identified promoter-distal regions with nucleosomes containing histone H3 globular domains that are acetylated at lysine residues 64 (H3K64Ac) and 122 (H3K122Ac) but not at H3K27 [409,410]. Enhancer-promoter contacts that do not overlap H3K27 acetylated regions may be indicative of H3K122 and H3K64 acetylated enhancers. They may also be indicative of negative regulators such as silencers, which have a repressive effect on the gene promoters they interact with and are also associated with a different epigenetic profile from enhancers marked with H3K27Ac [411,412]. Further studies to determine the function of these distinct regions will be required.

Differential enhancer usage, whereby two or more enhancers associated with the same gene undergo inversely related changes in H3K27Ac signal and promoter contact strength, was observed in enhancers associated with differentially expressed genes in leiomyomas. This mechanism may allow for the alternation of more densely occupied/ stronger enhancers with weaker ones and vice versa, resulting in changes in gene expression. It may also be a mechanism by which different groups of transcription factors and cofactors are brought into contact with promoters to modulate gene transcription in uterine leiomyomas.

Taken together, the observations of alterations in H3K27Ac signal at distal sites, altered promoter contacts, and the identification of differential enhancer usage in uterine leiomyomas, all suggest that in addition to important promoter-proximal gene transcription regulatory events, enhancer malfunction, a previously unexplored but significant area in leiomyoma disease pathogenesis, is an important mechanism of gene dysregulation in uterine leiomyomas.

While the loss of AP-1 complex subunit gene expression has been widely reported, the mechanisms by which this occurs and how AP-1 down regulation leads to changes in transcription are not clearly understood. This study describes the role of AP-1 in uterine muscle cell gene regulation and the mechanisms by which decreased AP-1 expression may contribute to the formation of uterine leiomyomas. A loss of *JUN* and *FOS* gene expression leads to a concomitant decrease in AP-1 recruitment at distal sites, which is positively correlated with changing H3K27Ac signal. Enhancers associated with differentially expressed genes were identified as also showing a change in AP-1 occupancy. These enhancers that predominantly had a decrease in AP-1 occupancy, were associated with both up regulated as well as down regulated genes. This may be indicative of the dual activating as well as repressive role associated with AP-1 and the possible existence and identification of other factors associated with these enhancers that may explain the dual role of AP-1 enhancers in leiomyomas will be important [339]. In addition, loss of AP-1 chromatin occupancy at enhancers partially overlaps with changes in promoter contact strength. This is consistent with the established role of AP-1 in mediating regulatory chromatin contacts and suggests that loss of AP-1 gene expression and decreased AP-1 chromatin occupancy may drive the remodelling of enhancer architecture in leiomyomas, resulting in the dysregulation of extracellular matrix associated genes [413,414].

Silencing of AP-1 subunits in human uterine smooth muscle cells (HUtSMCs) and myometrium (WT) primary cells, using CRISPR/Cas9 mediated gene editing and small hairpin-mediated knockdown respectively, confirms the role of AP-1 in enhancer malfunction-driven gene dysregulation. This work demonstrates that loss of AP-1 is sufficient to cause the large-scale changes seen in leiomyoma tissue enhancer acetylation as well as cause dysregulation of

many genes, including the perturbation of ECM genes, and thereby establishes a key role for AP-1 in uterine leiomyoma disease pathogenesis.

AP-1 depletion in HUtSMCs resulted in an altered H3K27Ac cistrome, which partially recapitulates H3K27Ac cistrome changes seen in leiomyoma tissue samples. Intergenic H3K27Ac is enriched at some sites and depleted at others in equal proportions, thereby demonstrating the major role played by AP-1 in enhancer maintenance in smooth muscle cells and how loss of AP-1 causes significant alterations to the cistrome. However, given that the loss of AP-1 does not primarily cause a concomitant depletion of H3K27Ac signal at intergenic sites, but rather results in a depletion of H3K27Ac at some sites and enrichment at other sites, further work will be required to determine if the increased H3K27Ac signal is due to indirect AP-1 effects. In addition, identification of the acetyltransferase and/ or deacetylase responsible for the alterations in H3K27Ac cistrome will be required. Similarly, work to determine if AP-1 is directly responsible for the recruitment of the H3K27-modifying enzymes will be equally important.

AP-1 is a transcription factor complex composed of immediate early genes and it regulates many signal-dependent processes in the cell. As such, it is a likely candidate for a master regulator of gene expression in uterine muscle cells, the loss of which may certainly explain the broad transcriptional changes seen in leiomyomas. Given the interest in AP-1 transcription factor activity at enhancers as well as the implication of AP-1 in multiple fibrotic diseases, future studies on AP-1 dependent enhancer function in other fibrotic diseases will be of great interest [414-417].

Down regulation of AP-1 is observed across leiomyomas of different subtypes [132]. Although this study identifies transcriptomic and epigenomic changes in *MED12* mutant fibroids, it is unclear if these changes are specific to *MED12* mutant leiomyomas or if the conclusions drawn from this study applies to leiomyomas in general. Given the role of AP-1 in widespread enhancer malfunction identified in this work, experiments to characterise the epigenomic and transcriptomic profiles of leiomyomas with genetic aberrations in *FH*, *HMGA2*, and type IV collagen genes should be carried out to determine if similar alterations in enhancer architecture are also observed in leiomyomas of non-*MED12* mutant origins, with AP-1 possibly providing a unifying mechanism of gene dysregulation in uterine leiomyomas of differing subtypes.

Given the overrepresentation of mutant *MED12* uterine leiomyomas in the patient population, *MED12* mutant tumours were used to carry out this study. However, the mechanism by which exon 2 mutations in *MED12* lead to leiomyomas is still not well understood. Studies have suggested that *MED12* mutations may affect transcription either through the disruption of CDK8 kinase activity or through decreased binding affinity of the MED12-containing CDK8 submodule at sites of active transcription [273]. This work demonstrates that whilst CDK8 submodule binding is altered in leiomyoma tissue samples, characterised by both a loss of binding affinity at a subset of sites and a recruitment of the CDK8 submodule at new sites in leiomyoma, the observed change is not consistent with a global loss of recruitment that would result from a mutation-dependent MED12–cyclin C–CDK8 interaction defect. Rather, these results show that changes in the CDK8 submodule cistrome are positively correlated with altered H3K27Ac, suggesting that altered submodule binding is more consistent with general changes in

the enhancer landscape of leiomyoma cells. This aligns with recently published data in heterologous cell lines as well as leiomyoma tissue samples that demonstrate that the presence of MED13 is a stabilising protein in the submodule that compensates for any MED12 mutant dependent interaction defect [274,418]. Therefore, *MED12* mutation-driven leiomyoma disease pathogenesis is more likely a result of CDK8 kinase activity defects.

Although multiple studies show decreased expression of AP-1 subunits in fibroids, a mechanism for AP-1 down regulation remains unknown. Despite a decrease in gene expression, AP-1 subunits do not show a change in H3K27Ac at gene promoters and although changes in a few enhancers associated with AP-1 family of genes are observed, it is unlikely that this is the primary mechanism by which AP-1 is dysregulated in uterine leiomyomas. The loss in expression of *JUN*, *JUNB*, *JUND*, *FOS*, *FOSB*, *FOSL1*, *FOSL2*, *ATF3*, *ATF6*, and *JDP2* would suggest a unifying mode of regulation for AP-1 family genes that is perturbed in uterine fibroids, however a singular enhancer or group of enhancers that jointly regulates AP-1 genes were not identified in this study.

Recently, Poss et al. demonstrated that although AP-1 subunits JUN and FOS are not CDK8 kinase substrates, their expression is down regulated upon treatment of cells with cortistatin A (CA), a potent and specific CDK8/CDK19 kinase inhibitor [419]. This reinforces the hypothesis that MED12 mutation-driven CDK8 kinase defects are important in leiomyoma disease development and are likely responsible for loss of AP-1 subunit expression. Future work will be required to identify the CDK8 kinase substrate that regulates AP-1 subunit expression and is no longer phosphorylated upon loss of CDK8 kinase activity in *MED12* mutant leiomyomas. In general, given the newly described role of AP-1 as a possible master regulator in

uterine muscle cells, work to further elucidate the mechanisms that result in the repression of AP-1 transcription factor subunit gene expression in leiomyomas will be of great significance in increasing our understanding of the mechanisms of uterine fibroid pathobiology.

Although overlap in chromatin occupancy of AP-1 and CDK8 complexes is demonstrated in this study, it is unclear if these two complexes directly interact and co-regulate transcription in uterine muscles cells. Protein-protein interaction studies will be required to determine if AP-1 – CDK8 submodule overlap is merely the co-localisation of two complexes on chromatin or if they do in fact directly interact. As JUN and FOS are not CDK8 kinase substrates, the significance of an AP-1 – CDK8 subcomplex direct interaction, if it indeed exists, is unclear but may indicate a structural role of the two complexes in the stabilisation of enhancer-promoter contacts.

CONCLUSION AND FUTURE DIRECTIONS

In summary, this work highlights enhancer malfunction as an important and previously unexplored mechanism of gene dysregulation in *MED12* mutant uterine leiomyomas, including the dysregulation of genes involved in key biological processes in the disease such as extracellular matrix organisation. Altered enhancer architecture, which consists of changes in histone acetylation and alterations in promoter contact strength, is driven in part by a loss of AP-1 subunit expression, resulting in depleted AP-1 transcription factor occupancy at enhancers.

In light of previously reported perturbations in AP-1 gene regulation and signalling in leiomyomas of a non-*MED12* mutant subtype, as well as the reported changes in AP-1 signalling in other fibrotic diseases, it is possible that AP-1 dysregulation is a common mechanism of pathogenesis in fibrotic diseases. As such, further work to describe AP-1 signalling in fibroids of a non-*MED12* mutant subtype will be important. Characterisation of the epigenomic status of non-*MED12* mutant leiomyomas will be necessary to determine if similarities in enhancer architecture exist between *MED12* mutant and none *MED12* mutant leiomyomas.

Despite evidence presented in this study detailing the role of AP-1 in altered leiomyoma enhancer architecture and gene dysregulation, it remains to be experimentally determined if the loss of AP-1 is either necessary or sufficient for the development of uterine leiomyomas. Also, although all *JUN* and *FOS* family genes are down regulated in uterine fibroids, it is unclear if the loss of expression of all of these genes is necessary to cause the enhancer defects observed in leiomyoma cells. Given that incomplete CRISPR/Cas9 mediated depletion of some AP-1 family member genes in this study still results in enhancer malfunction, it is likely that a subset of these genes are responsible for the observed dysfunction. As AP-1 is a dimeric complex whose composition determines its regulatory function, combinatorial depletion of various *FOS* and *JUN*

family genes could be responsible for regulating different subsets of enhancers and genes. Further studies will be necessary to ascertain if AP-1 is required for leiomyoma pathogenesis and which subunits in particular are important in this disease.

Previous studies suggest that the mutation of *MED12* in uterine leiomyomas leads to a CDK8 kinase defect. However, similarly to AP-1, it remains undetermined if the *MED12* mutation-driven CDK8 kinase defect observed in leiomyomas is a causal defect in the disease. This is a key aspect of the mutant MED12-CDK8 mechanistic pathway that needs to be experimentally tested and a mouse model expressing a uterine tissue-specific CDK8 kinase-dead mutant may be an appropriate vehicle for this purpose.

While previous studies do not suggest that the CDK8 subcomplex directly phosphorylates AP-1 family genes, it is possible that a kinase substrate of CDK8 serves as a central regulator of AP-1 genes. Alternatively, Mediator kinases have been demonstrated to primarily "regulate the regulators" of transcription and have also been implicated in RNAPII pause-release [419]. As AP-1 is a complex of immediate early genes, it is conceivable that the CDK8 subcomplex regulates AP-1 genes through RNAPII pause release. Given the varied possibilities regarding the role of a CDK8 kinase defect in *MED12* mutant leiomyoma cells, future work to unambiguously define the mechanism of CDK8 action in these cells will be of great importance.

This work establishes significant evidence for the interplay between CDK8 subcomplex kinase activity and AP-1 dependent gene regulation in uterine leiomyomas. Future studies should focus on delineating the mechanism of CDK8 mediated AP-1 gene regulation and how the perturbation of this CDK8 submodule – AP-1 complex interplay results in enhancer malfunction in uterine leiomyomas.

APPENDIX

Table 5.1. Clinical and biological characteristics of leiomyomas obtained from hysterectomies of 15 women.

Patient # in figures	De-identified Patient # in sequencing data	Mutation	Age	Race/ Ethnicity	Hormone treatment
PT1	PT848	G44S	49	Black	Progesterone
PT2	PT886	G44D	42	White	None
PT3	PT916	G44D	51	White	Lo-estrin
PT4	PT967	G44D	43	Unknown	None
PT5	PT1063	G44D	43	Black	None
PT6	PT354	G44D	38	Black	None
PT7	PT563	G44S	45	Black	None
PT8	PT728	G44D	44	Latino	None
PT9	PT758	G44D	48	Black	None
PT10	PT845	G44D	48	Black	None
PT11	PT1113	G44D	51	Unknown	None
PT12	PT1119	G44S	44	White	None
PT13	PT1123	G44S	48	Unknown	None
PT14	PT1151	G44S	47	Black	None
PT15	PTC57	G44D	50	Black	None

5.2. qRT-PCR primers

*JUN*_F: 5´-GTCCTTCTTCTCTTGCGTGG-3´
*JUN*_R: 5´-GGAGACAAGTGGCAGAGTCC-3´
*JUNB*_F: 5´-AGGCTCGGTTTCAGGAGTTT-3´
*JUNB*_R: 5´-GAACAGCCCTTCTACCACGA-3´
*JUND*_F: 5´-CACCTTGGGGTAGAGGAACTG-3´
*JUND*_R: 5´-CCTCATCATCCAGTCCAACGG-3´
*FOS*_F: 5´-CTACCACTCACCCGCAGACT-3´
*FOS*_R: 5´-GTGGGAATGAAGTTGGCACT-3´
*FOSB*_F: 5´-CTAGGAGACCCCGAGAGGAG-3´
*FOSB*_R: 5´-ACCAGCACAAACTCCAGACG-3´

5.3. Short hairpin target sequences

JUN (TRCN0000039590): 5´-CGCAAACCTCAGCAACTTCAA-3´
JUNB (TRCN0000232083): 5´-TCATACACAGCTACGGGATAC-3´
JUND (TRCN0000416920): 5´-GAAGAACAGAGTGTTCGATTC-3´
FOS (TRCN0000273940): 5´-TCTCCAGTGCCAACTTCATTC-3´
FOSB (TRCN0000016071): 5´-GCCGAGTCTCAATATCTGTCT-3´

5.4. CRISPR/Cas9 crRNA target sequences

JUN (Hs.Cas9.JUN.1.AA): CCATAAGGTCCGCTCTCGGA
JUNB (Hs.Cas9.JUNB.1.AB): CGACGACTCATACACAGCTA
JUND (Hs.Cas9.JUND.1.AB): TGGTGACCAGCCCGTTGGAC
FOS (Hs.Cas9.FOS.1.AA): GGGCTTCAACGCAGACTACG
FOSB (Hs.Cas9.FOSB.1.AA): CGTCGACCCCTACGACATGC

5.5. gDNA amplification primers

*JUN*_F: 5´-TCCTGGGACTCCATGTCGAT-3´
*JUN*_R: 5´-TGCGTGCGCTCTTAGAGAAA-3´
*JUNB*_F: 5´-AAACGACGCCAGGAAAGCTA-3´
*JUNB*_R: 5´-GGGTGTCACGTGGTTCATCT-3´
*JUND*_F: 5´-CTCAGGTTCGCGTAGACAGG-3´ (Requires DMSO and betadine for PCR)
*JUND*_R: 5´-CTCCATGCAAATGAGCGACG-3´ (Requires DMSO and betadine for PCR)
*FOS*_F: 5´-CAGTTCCCGTCAATCCCTCC-3´
*FOS*_R: 5´-TCTGCGGGTGAGTGGTAGTA-3´
*FOSB*_F: 5´-TCTCCTTCTCCCTCCCCTTG-3´
*FOSB*_R: 5´-CCTACCCAGGCCAAGTTCTG-3´
*MED12*_F (exon 2): 5´-GCTGGGAATCCTAGTGACCA-3´
*MED12*_R (exon 2): 5´-GGCAAACTCAGCCACTTAGG-3´

REFERENCES

1. Stewart, E. A. *et al.* Uterine fibroids. *Nat Rev Dis Primers* **2,** 16043 (2016).

2. Moroni, R. M. *et al.* Presentation and treatment of uterine leiomyoma in adolescence: a systematic review. *BMC Womens Health* **15,** 4 (2015).

3. Fields, K. R. & Neinstein, L. S. Uterine myomas in adolescents: case reports and a review of the literature. *J Pediatr Adolesc Gynecol* **9,** 195–198 (1996).

4. Wang, C. P., Chang, Y. L. & Sheen, T. S. Vascular leiomyoma of the head and neck. *Laryngoscope* **114,** 661–665 (2004).

5. Tomlinson, I. P. M. *et al.* Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* **30,** 406–410 (2002).

6. Punpale, A. *et al.* Leiomyoma of esophagus. *Ann Thorac Cardiovasc Surg* **13,** 78–81 (2007).

7. Llorente, J. Laparoscopic gastric resection for gastric leiomyoma. *Surg Endosc* **8,** 887–889 (1994).

8. Mutrie, C. J. *et al.* Esophageal leiomyoma: a 40-year experience. *Ann. Thorac. Surg.* **79,** 1122–1125 (2005).

9. Park, J. W. *et al.* Leiomyoma of the urinary bladder: a series of nine cases and review of the literature. *Urology* **76,** 1425–1429 (2010).

10. Ricci, S., Stone, R. L. & Fader, A. N. Uterine leiomyosarcoma: Epidemiology, contemporary treatment strategies and the impact of uterine morcellation. *Gynecol. Oncol.* **145,** 208–216 (2017).

11. D'Angelo, E. & Prat, J. Uterine sarcomas: a review. *Gynecol. Oncol.* **116,** 131–139 (2010).

12. Ravegnini, G. *et al.* MED12 mutations in leiomyosarcoma and extrauterine leiomyoma. *Mod Pathol* **26,** 743–749 (2013).

13. Munro, M. G., Critchley, H. O. D., Broder, M. S., Fraser, I. S.FIGO Working Group on Menstrual Disorders. FIGO classification system (PALM-COEIN) for causes of abnormal uterine bleeding in nongravid women of reproductive age. *Int J Gynaecol Obstet* **113,** 3–13 (2011).

14. Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D. & Schectman, J. M. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. *Am. J. Obstet. Gynecol.* **188,** 100–107 (2003).

15. Marshall, L. M. *et al.* Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. *Obstet Gynecol* **90,** 967–973 (1997).

16. Drayer, S. M. & Catherino, W. H. Prevalence, morbidity, and current medical management of uterine leiomyomas. *Int J Gynaecol Obstet* **131,** 117–122 (2015).

17. Peddada, S. D. *et al.* Growth of uterine leiomyomata among premenopausal black and white women. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 19887–19892 (2008).

18. Cramer, S. F. & Patel, A. The frequency of uterine leiomyomas. *Am. J. Clin. Pathol.* **94,** 435–438 (1990).

19. Marsh, E. E. *et al.* Racial differences in fibroid prevalence and ultrasound findings in asymptomatic young women (18-30 years old): a pilot study. *Fertil. Steril.* **99,** 1951–1957 (2013).

20. Marshall, L. M. *et al.* Variation in the incidence of uterine leiomyoma among

premenopausal women by age and race. *Obstet Gynecol* **90,** 967–973 (1997).

21. DeWaay, D. J., Syrop, C. H., Nygaard, I. E., Davis, W. A. & Van Voorhis, B. J. Natural history of uterine polyps and leiomyomata. *Obstet Gynecol* **100,** 3–7 (2002).

22. BORGFELDT, C. & ANDOLF, E. Transvaginal ultrasonographic findings in the uterus and the endometrium: Low prevalence of leiomyoma in a random sample of women age 25-40 years. *Acta Obstet Gynecol Scand* **79,** 202–207 (2000).

23. Parker, W. H. Etiology, symptomatology, and diagnosis of uterine myomas. *Fertil. Steril.* **87,** 725–736 (2007).

24. Lippman, S. A. *et al.* Uterine fibroids and gynecologic pain symptoms in a population-based study. *Fertil. Steril.* **80,** 1488–1494 (2003).

25. Langer, R. *et al.* The effect of large uterine fibroids on urinary bladder function and symptoms. *Am. J. Obstet. Gynecol.* **163,** 1139–1141 (1990).

26. Pritts, E. A., Parker, W. H. & Olive, D. L. Fibroids and infertility: an updated systematic review of the evidence. *Fertil. Steril.* **91,** 1215–1223 (2009).

27. Klatsky, P. C., Tran, N. D., Caughey, A. B. & Fujimoto, V. Y. Fibroids and reproductive outcomes: a systematic literature review from conception to delivery. *Am. J. Obstet. Gynecol.* **198,** 357–366 (2008).

28. Somigliana, E. *et al.* Fibroids and female reproduction: a critical analysis of the evidence. *Hum. Reprod. Update* **13,** 465–476 (2007).

29. Richards, P. A., Richards, P. D. & Tiltman, A. J. The ultrastructure of fibromyomatous myometrium and its relationship to infertility. *Hum. Reprod. Update* **4,** 520–525 (1998).

30. Whynott, R. M., Vaught, K. C. C. & Segars, J. H. The Effect of Uterine Fibroids on

Infertility: A Systematic Review. *Semin. Reprod. Med.* **35,** 523–532 (2017).

31. Levens, E. D. *et al.* CDB-2914 for uterine leiomyomata treatment: a randomized controlled trial. *Obstet Gynecol* **111,** 1129–1136 (2008).

32. Chabbert-Buffet, N., Meduri, G., Bouchard, P. & Spitz, I. M. Selective progesterone receptor modulators and progesterone antagonists: mechanisms of action and clinical applications. *Hum. Reprod. Update* **11,** 293–307 (2005).

33. Talaulikar, V. S. & Manyonda, I. Progesterone and progesterone receptor modulators in the management of symptomatic uterine fibroids. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **165,** 135–140 (2012).

34. Donnez, J. *et al.* Long-term medical management of uterine fibroids with ulipristal acetate. *Fertil. Steril.* **105,** 165–173.e4 (2016).

35. Donnez, J. *et al.* Ulipristal acetate versus placebo for fibroid treatment before surgery. *N Engl J Med* **366,** 409–420 (2012).

36. Williams, A. R. W. *et al.* The effects of the selective progesterone receptor modulator asoprisnil on the morphology of uterine tissues after 3 months treatment in patients with symptomatic uterine leiomyomata. *Hum. Reprod.* **22,** 1696–1704 (2007).

37. Stewart, E. A. *et al.* Safety and efficacy of the selective progesterone receptor modulator asoprisnil for heavy menstrual bleeding with uterine fibroids: pooled analysis of two 12-month, placebo-controlled, randomized trials. *Hum. Reprod.* **34,** 623–634 (2019).

38. Sayed, G. H., Zakherah, M. S., Nashar, El, S. A. & Shaaban, M. M. A randomized clinical trial of a levonorgestrel-releasing intrauterine system and a low-dose combined oral contraceptive for fibroid-related menorrhagia. *Int J Gynaecol Obstet* **112,** 126–130 (2011).

39.   Murat Naki, M., Tekcan, C., Ozcan, N. & Cebi, M. Levonorgestrel-releasing intrauterine
       device insertion ameliorates leiomyoma-dependent menorrhagia among women of
       reproductive age without a significant regression in the uterine and leiomyoma volumes.
       Fertil. Steril. **94,** 371–374 (2010).

40.   Wildemeersch, D. & Schacht, E. The effect on menstrual blood loss in women with
       uterine fibroids of a novel 'frameless' intrauterine levonorgestrel-releasing drug delivery
       system: a pilot study. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **102,** 74–79 (2002).

41.   Sriprasert, I., Pakrashi, T., Kimble, T. & Archer, D. F. Heavy menstrual bleeding
       diagnosis and medical management. *Contracept Reprod Med* **2,** 20 (2017).

42.   Guttinger, A. & Critchley, H. O. D. Endometrial effects of intrauterine levonorgestrel.
       Contraception **75,** S93–8 (2007).

43.   Maybin, J. A. & Critchley, H. O. D. Medical management of heavy menstrual bleeding.
       Womens Health (Lond) **12,** 27–34 (2016).

44.   Walker, C. L. *et al.* Preclinical evidence for therapeutic efficacy of selective estrogen
       receptor modulators for uterine leiomyoma. *J. Soc. Gynecol. Investig.* **7,** 249–256 (2000).

45.   Young, R. F., Howe, S., Hale, L., Miles, R. & Walker, C. Inhibition of estrogen-stimulated
       growth of uterine leiomyomas by selective estrogen receptor modulators. *Molecular
       Carcinogenesis* **17,** 151–159 (1996).

46.   Jirecek, S. *et al.* Raloxifene prevents the growth of uterine leiomyomas in premenopausal
       women. *Fertil. Steril.* **81,** 132–136 (2004).

47.   Palomba, S. *et al.* Antiproliferative and proapoptotic effects of raloxifene on uterine
       leiomyomas in postmenopausal women. *Fertil. Steril.* **84,** 154–161 (2005).

48.   Simpson, E. R. *et al.* Aromatase--a brief overview. *Annu. Rev. Physiol.* **64,** 93–127 (2002).

49.   Akhtar, M., Wright, J. N. & Lee-Robichaud, P. A review of mechanistic studies on aromatase (CYP19) and 17α-hydroxylase-17,20-lyase (CYP17). *The Journal of Steroid Biochemistry and Molecular Biology* **125,** 2–12 (2011).

50.   Varelas, F. K., Papanicolaou, A. N., Vavatsi-Christaki, N., Makedos, G. A. & Vlassis, G. D. The effect of anastrazole on symptomatic uterine leiomyomata. *Obstet Gynecol* **110,** 643–649 (2007).

51.   Hilário, S. G., Bozzini, N., Borsari, R. & Baracat, E. C. Action of aromatase inhibitor for treatment of uterine leiomyoma in perimenopausal patients. *Fertil. Steril.* **91,** 240–243 (2009).

52.   Parsanezhad, M. E. *et al.* A randomized, controlled clinical trial comparing the effects of aromatase inhibitor (letrozole) and gonadotropin-releasing hormone agonist (triptorelin) on uterine leiomyoma volume and hormonal status. *Fertil. Steril.* **93,** 192–198 (2010).

53.   Duhan, N., Madaan, S. & Sen, J. Role of the aromatase inhibitor letrozole in the management of uterine leiomyomas in premenopausal women. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **171,** 329–332 (2013).

54.   Thackray, V. G., Mellon, P. L. & Coss, D. Hormones in synergy: regulation of the pituitary gonadotropin genes. *Molecular and Cellular Endocrinology* **314,** 192–203 (2010).

55.   Herbison, A. E. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. *Nat Rev Endocrinol* **12,** 452–466 (2016).

56.   Wilson, A. C., Meethal, S. V., Bowen, R. L. & Atwood, C. S. Leuprolide acetate: a drug

of diverse clinical applications. *Expert Opin Investig Drugs* **16,** 1851–1863 (2007).

57. Friedman, A. J., Hoffman, D. I., Comite, F., Browneller, R. W. & Miller, J. D. Treatment of leiomyomata uteri with leuprolide acetate depot: a double-blind, placebo-controlled, multicenter study. The Leuprolide Study Group. *Obstet Gynecol* **77,** 720–725 (1991).

58. Huirne, J. A. & Lambalk, C. B. Gonadotropin-releasing-hormone-receptor antagonists. *The Lancet* **358,** 1793–1803 (2001).

59. Felberbaum, R. E. *et al.* Treatment of uterine fibroids with a slow-release formulation of the gonadotrophin releasing hormone antagonist Cetrorelix. *Hum. Reprod.* **13,** 1660–1668 (1998).

60. Flierman, P. A., Oberyé, J. J. L., van der Hulst, V. P. M. & de Blok, S. Rapid reduction of leiomyoma volume during treatment with the GnRH antagonist ganirelix. *BJOG* **112,** 638–642 (2005).

61. Archer, D. F. *et al.* Elagolix for the management of heavy menstrual bleeding associated with uterine fibroids: results from a phase 2a proof-of-concept study. *Fertil. Steril.* **108,** 152–160.e4 (2017).

62. Malik, M., Britten, J., Cox, J., Patel, A. & Catherino, W. H. Gonadotropin-releasing hormone analogues inhibit leiomyoma extracellular matrix despite presence of gonadal hormones. *Fertil. Steril.* **105,** 214–224 (2016).

63. Britten, J. L., Malik, M., Levy, G., Mendoza, M. & Catherino, W. H. Gonadotropin-releasing hormone (GnRH) agonist leuprolide acetate and GnRH antagonist cetrorelix acetate directly inhibit leiomyoma extracellular matrix production. *Fertil. Steril.* **98,** 1299–1307 (2012).

64. Wechter, M. E., Stewart, E. A., Myers, E. R., Kho, R. M. & Wu, J. M. Leiomyoma-related hospitalization and surgery: prevalence and predicted growth based on population trends. *Am. J. Obstet. Gynecol.* **205,** 492.e1–5 (2011).

65. Wright, J. D. *et al.* Nationwide Trends in the Performance of Inpatient Hysterectomy in the United States. *Obstet Gynecol* **122,** 233–241 (2013).

66. Samejima, T. *et al.* Identifying patients who can improve fertility with myomectomy. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **185,** 28–32 (2015).

67. Li, T. C., Mortimer, R. & Cooke, I. D. Myomectomy: a retrospective study to examine reproductive performance before and after surgery. *Hum. Reprod.* **14,** 1735–1740 (1999).

68. Di Spiezio Sardo, A. *et al.* Hysteroscopic myomectomy: a comprehensive review of surgical techniques. *Hum. Reprod. Update* **14,** 101–119 (2008).

69. Kotani, Y. *et al.* Recurrence of uterine myoma after myomectomy: Open myomectomy versus laparoscopic myomectomy. *J. Obstet. Gynaecol. Res.* **44,** 298–302 (2018).

70. Buckley, V. A. *et al.* Laparoscopic myomectomy: clinical outcomes and comparative evidence. *J Minim Invasive Gynecol* **22,** 11–25 (2015).

71. Parker, W. H. Laparoscopic myomectomy and abdominal myomectomy. *Clin Obstet Gynecol* **49,** 789–797 (2006).

72. Fennessy, F. M. & Tempany, C. M. MRI-guided focused ultrasound surgery of uterine leiomyomas. *Academic Radiology* **12,** 1158–1166 (2005).

73. Rabinovici, J. *et al.* Pregnancy outcome after magnetic resonance-guided focused ultrasound surgery (MRgFUS) for conservative treatment of uterine fibroids. *Fertil. Steril.* **93,** 199–209 (2010).

74. Rabinovici, J. *et al.* Clinical improvement and shrinkage of uterine fibroids after thermal ablation by magnetic resonance-guided focused ultrasound surgery. *Ultrasound Obstet Gynecol* **30,** 771–777 (2007).

75. Stewart, E. A. *et al.* Sustained relief of leiomyoma symptoms by using focused ultrasound surgery. *Obstet Gynecol* **110,** 279–287 (2007).

76. de Bruijn, A. M. *et al.* Uterine artery embolization vs hysterectomy in the treatment of symptomatic uterine fibroids: 10-year outcomes from the randomized EMMY trial. *Am. J. Obstet. Gynecol.* **215,** 745.e1–745.e12 (2016).

77. Spies, J. B. *et al.* Initial Results from Uterine Fibroid Embolization for Symptomatic Leiomyomata. *Journal of Vascular and Interventional Radiology* **10,** 1149–1157 (1999).

78. Mara, M., Fucikova, Z., Maskova, J., Kuzel, D. & Haakova, L. Uterine fibroid embolization versus myomectomy in women wishing to preserve fertility: preliminary results of a randomized controlled trial. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **126,** 226–233 (2006).

79. Stewart, E. A. Clinical practice. Uterine fibroids. *N Engl J Med* **372,** 1646–1655 (2015).

80. Cardozo, E. R. *et al.* The estimated annual cost of uterine leiomyomata in the United States. *Am. J. Obstet. Gynecol.* **206,** 211.e1–9 (2012).

81. Linder, D. & Gartler, S. M. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science* **150,** 67–69 (1965).

82. Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.*

51, 1229–1239 (1992).

83. Mashal, R. D. *et al.* Analysis of androgen receptor DNA reveals the independent clonal origins of uterine leiomyomata and the secondary nature of cytogenetic aberrations in the development of leiomyomata. *Genes, Chromosomes and Cancer* **11,** 1–6 (1994).

84. Cai, Y.-R. *et al.* X-chromosomal inactivation analysis of uterine leiomyomas reveals a common clonal origin of different tumor nodules in some multiple leiomyomas. *Int. J. Oncol.* **31,** 1379–1389 (2007).

85. Zhang, P. *et al.* Use of X-chromosome inactivation pattern to determine the clonal origins of uterine leiomyoma and leiomyosarcoma. *Human Pathology* **37,** 1350–1356 (2006).

86. Holdsworth-Carson, S. J., Zaitseva, M., Vollenhoven, B. J. & Rogers, P. A. W. Clonality of smooth muscle and fibroblast cell populations isolated from human fibroid and myometrial tissues. *Mol. Hum. Reprod.* **20,** 250–259 (2014).

87. Mas, A. *et al.* Identification and characterization of the human leiomyoma side population as putative tumor-initiating cells. *Fertil. Steril.* **98,** 741–751.e6 (2012).

88. Ono, M. *et al.* Role of stem cells in human uterine leiomyoma growth. *PLoS ONE* **7,** e36935 (2012).

89. Dixon, D. *et al.* Cell proliferation and apoptosis in human uterine leiomyomas and myometria. *Virchows Arch.* **441,** 53–62 (2002).

90. Malik, M., Norian, J., McCarthy-Keith, D., Britten, J. & Catherino, W. H. Why leiomyomas are called fibroids: the central role of extracellular matrix in symptomatic women. *Semin. Reprod. Med.* **28,** 169–179 (2010).

91. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell.*

Sci. **123,** 4195–4200 (2010).

92. Rozario, T. & DeSimone, D. W. The extracellular matrix in development and morphogenesis: a dynamic view. *Dev. Biol.* **341,** 126–140 (2010).

93. Hynes, R. O. The extracellular matrix: not just pretty fibrils. *Science* **326,** 1216–1219 (2009).

94. Lu, P., Takai, K., Weaver, V. M. & Werb, Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb Perspect Biol* **3,** (2011).

95. Herrera, J., Henke, C. A. & Bitterman, P. B. Extracellular matrix as a driver of progressive fibrosis. *J. Clin. Invest.* **128,** 45–53 (2018).

96. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *The Journal of Cell Biology* **196,** 395–406 (2012).

97. Fan, D., Takawale, A., Lee, J. & Kassiri, Z. Cardiac fibroblasts, fibrosis and extracellular matrix remodeling in heart disease. *Fibrogenesis Tissue Repair* **5,** 15 (2012).

98. Bedossa, P. & Paradis, V. Liver extracellular matrix in health and disease. *J. Pathol.* **200,** 504–515 (2003).

99. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* **15,** 786–801 (2014).

100. Muschler, J. & Streuli, C. H. Cell-matrix interactions in mammary gland development and breast cancer. *Cold Spring Harb Perspect Biol* **2,** a003202–a003202 (2010).

101. Theocharis, A. D., Skandalis, S. S., Tzanakakis, G. N. & Karamanos, N. K. Proteoglycans in health and disease: novel roles for proteoglycans in malignancy and their pharmacological targeting. *FEBS J.* **277,** 3904–3923 (2010).

102. Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. Collagens at a glance. *J. Cell. Sci.* **120,** 1955–1958 (2007).

103. Theocharis, A. D., Skandalis, S. S., Gialeli, C. & Karamanos, N. K. Extracellular matrix structure. *Adv. Drug Deliv. Rev.* **97,** 4–27 (2016).

104. Leppert, P. C., Catherino, W. H. & Segars, J. H. A new hypothesis about the origin of uterine fibroids based on gene expression profiling with microarrays. *Am. J. Obstet. Gynecol.* **195,** 415–420 (2006).

105. Wolańska, M., Sobolewski, K., Drozdzewicz, M. & Bańkowski, E. Extracellular matrix components in uterine leiomyoma and their alteration during the tumour growth. *Mol. Cell. Biochem.* **189,** 145–152 (1998).

106. Pankov, R. & Yamada, K. M. Fibronectin at a glance. *J. Cell. Sci.* **115,** 3861–3863 (2002).

107. Arici, A. & Sozen, I. Transforming growth factor-beta3 is expressed at high levels in leiomyoma where it stimulates fibronectin expression and cell proliferation. *Fertil. Steril.* **73,** 1006–1011 (2000).

108. Debelle, L. & Alix, A. J. The structures of elastins and their function. *Biochimie* **81,** 981–994 (1999).

109. Kielty, C. M., Sherratt, M. J. & Shuttleworth, C. A. Elastic fibres. *J. Cell. Sci.* **115,** 2817–2828 (2002).

110. Colombatti, A. *et al.* The EMILIN protein family. *Matrix Biology* **19,** 289–301 (2000).

111. Wolańska, M., Sobolewski, K., Cechowska-Pasko, M. & Jaworski, S. The activities of some glycosaminoglycan-degrading enzymes in uterine leiomyomas. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **110,** 73–78 (2003).

112. Colognato, H. & Yurchenco, P. D. Form and function: the laminin family of heterotrimers. *Dev. Dyn.* **218,** 213–234 (2000).

113. Malik, M., Segars, J. & Catherino, W. H. Integrin β1 regulates leiomyoma cytoskeletal integrity and growth. *Matrix Biol.* **31,** 389–397 (2012).

114. Prydz, K. & Dalen, K. T. Synthesis and sorting of proteoglycans. *J. Cell. Sci.* **113 Pt 2,** 193–205 (2000).

115. Wight, T. N. & Merrilees, M. J. Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circ. Res.* **94,** 1158–1167 (2004).

116. Iozzo, R. V. & Sanderson, R. D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J. Cell. Mol. Med.* **15,** 1013–1031 (2011).

117. Krishnan, A. *et al.* Lumican, an extracellular matrix proteoglycan, is a novel requisite for hepatic fibrosis. *Lab. Invest.* **92,** 1712–1725 (2012).

118. Kolb, M., Margetts, P. J., Sime, P. J. & Gauldie, J. Proteoglycans decorin and biglycan differentially modulate TGF-beta-mediated fibrotic responses in the lung. *Am. J. Physiol. Lung Cell Mol. Physiol.* **280,** L1327–34 (2001).

119. Schaefer, L. Small leucine-rich proteoglycans in kidney disease. *J. Am. Soc. Nephrol.* **22,** 1200–1207 (2011).

120. Mitropoulou, T. N., Theocharis, A. D., Stagiannis, K. D. & Karamanos, N. K. Identification, quantification and fine structural characterization of glycosaminoglycans from uterine leiomyoma and normal myometrium. *Biochimie* **83,** 529–536 (2001).

121. Barker, N. M. *et al.* Proteoglycans in Leiomyoma and Normal Myometrium: Abundance, Steroid Hormone Control, and Implications for Pathophysiology. *Reprod Sci* **23,** 302–309

(2016).

122. Islam, M. S., Ciavattini, A., Petraglia, F., Castellucci, M. & Ciarmela, P. Extracellular matrix in uterine leiomyoma pathogenesis: a potential target for future therapeutics. *Hum. Reprod. Update* **24,** 59–85 (2018).

123. Hynes, R. O. & Naba, A. Overview of the Matrisome—An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harb Perspect Biol* **4,** a004903–a004903 (2012).

124. Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell Proteomics* **11,** M111.014647 (2012).

125. Trackman, P. C. Enzymatic and non-enzymatic functions of the lysyl oxidase family in bone. *Matrix Biol.* **52-54,** 7–18 (2016).

126. Taipale, J. & Keski-Oja, J. Growth factors in the extracellular matrix. *FASEB J.* **11,** 51–59 (1997).

127. Pusztai, L., Lewis, C. E., Lorenzen, J. & McGee, J. O. Growth factors: regulation of normal and neoplastic growth. *J. Pathol.* **169,** 191–201 (1993).

128. Ciarmela, P. *et al.* Growth factors and myometrium: biological effects in uterine fibroid and possible clinical implications. *Hum. Reprod. Update* **17,** 772–790 (2011).

129. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2,** 127–137 (2001).

130. Werner, S. & Grose, R. Regulation of wound healing by growth factors and cytokines. *Physiol. Rev.* **83,** 835–870 (2003).

131. Linggi, B. & Carpenter, G. ErbB receptors: new insights on mechanisms and biology. *Trends in Cell Biology* **16,** 649–656 (2006).

132. Mehine, M. *et al.* Integrated data analysis reveals uterine leiomyoma subtypes with distinct driver pathways and biomarkers. *Proceedings of the National Academy of Sciences* **113,** 1315–1320 (2016).

133. Dixon, D., He, H. & Haseman, J. K. Immunohistochemical localization of growth factors and their receptors in uterine leiomyomas and matched myometrium. *Environ. Health Perspect.* **108 Suppl 5,** 795–802 (2000).

134. Wang, J., Ohara, N., Takekida, S., Xu, Q. & Maruo, T. Comparative effects of heparin-binding epidermal growth factor-like growth factor on the growth of cultured human uterine leiomyoma cells and myometrial cells. *Hum. Reprod.* **20,** 1456–1465 (2005).

135. Shushan, A. *et al.* The AG1478 tyrosine kinase inhibitor is an effective suppressor of leiomyoma cell growth. *Hum. Reprod.* **19,** 1957–1967 (2004).

136. Shushan, A. *et al.* Inhibition of leiomyoma cell proliferation in vitro by genistein and the protein tyrosine kinase inhibitor TKS050. *Fertil. Steril.* **87,** 127–135 (2007).

137. Zhang, X. *et al.* Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *J. Biol. Chem.* **281,** 15694–15700 (2006).

138. Turner, N. & Grose, R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer* **10,** 116–129 (2010).

139. Wolańska, M. & Bańkowski, E. Fibroblast growth factors (FGF) in human myometrium and uterine leiomyomas in various stages of tumour growth. *Biochimie* **88,** 141–146 (2006).

140. Nowak, R. A. Novel therapeutic strategies for leiomyomas: targeting growth factors and their receptors. *Environ. Health Perspect.* **108 Suppl 5,** 849–853 (2000).

141. Anania, C. A., Stewart, E. A., Quade, B. J., Hill, J. A. & Nowak, R. A. Expression of the fibroblast growth factor receptor in women with leiomyomas and abnormal uterine bleeding. *Mol. Hum. Reprod.* **3,** 685–691 (1997).

142. Pollak, M. Insulin and insulin-like growth factor signalling in neoplasia. *Nat Rev Cancer* **8,** 915–928 (2008).

143. Firth, S. M. & Baxter, R. C. Cellular actions of the insulin-like growth factor binding proteins. *Endocr. Rev.* **23,** 824–854 (2002).

144. Fowlkes, J. L., Enghild, J. J., Suzuki, K. & Nagase, H. Matrix metalloproteinases degrade insulin-like growth factor-binding protein-3 in dermal fibroblast cultures. *J. Biol. Chem.* **269,** 25742–25746 (1994).

145. Cirilo, P. D. R. *et al.* An Integrative Genomic and Transcriptomic Analysis Reveals Potential Targets Associated with Cell Proliferation in Uterine Leiomyomas. *PLoS ONE* **8,** e57901 (2013).

146. Peng, L. *et al.* Expression of insulin-like growth factors (IGFs) and IGF signaling: molecular complexity in uterine leiomyomas. *Fertil. Steril.* **91,** 2664–2675 (2009).

147. Andrae, J., Gallini, R. & Betsholtz, C. Role of platelet-derived growth factors in physiology and medicine. *Genes & Development* **22,** 1276–1312 (2008).

148. Matsumoto, T. & Claesson-Welsh, L. VEGF receptor signal transduction. *Sci. STKE* **2001,** re21–re21 (2001).

149. Ferrara, N. Vascular endothelial growth factor: basic science and clinical progress.

Endocr. Rev. **25,** 581–611 (2004).

150. Ball, S. G., Shuttleworth, C. A. & Kielty, C. M. Vascular endothelial growth factor can signal through platelet-derived growth factor receptors. *The Journal of Cell Biology* **177,** 489–500 (2007).

151. Ferrara, N., Gerber, H.-P. & LeCouter, J. The biology of VEGF and its receptors. *Nat. Med.* **9,** 669–676 (2003).

152. Boor, P. & Floege, J. Chronic kidney disease growth factors in renal fibrosis. *Clin. Exp. Pharmacol. Physiol.* **38,** 441–450 (2011).

153. Chaudhary, N. I. *et al.* Inhibition of PDGF, VEGF and FGF signalling attenuates fibrosis. *Eur. Respir. J.* **29,** 976–985 (2007).

154. Liang, M., Wang, H., Zhang, Y., Lu, S. & Wang, Z. Expression and functional analysis of platelet-derived growth factor in uterine leiomyomata. *Cancer Biol. Ther.* **5,** 28–33 (2006).

155. Hwu, Y.-M. *et al.* Increased expression of platelet-derived growth factor C messenger ribonucleic acid in uterine leiomyomata. *Fertil. Steril.* **89,** 468–471 (2008).

156. Di Lieto, A. *et al.* Preoperative administration of GnRH-a plus tibolone to premenopausal women with uterine fibroids: evaluation of the clinical response, the immunohistochemical expression of PDGF, bFGF and VEGF and the vascular pattern. *Steroids* **70,** 95–102 (2005).

157. Taniguchi, Y., Morita, I., Kubota, T., Murota, S. & Aso, T. Human uterine myometrial smooth muscle cell proliferation and vascular endothelial growth-factor production in response to platelet-derived growth factor. *J. Endocrinol.* **169,** 79–86 (2001).

158. Massagué, J. TGFβ signalling in context. *Nat Rev Mol Cell Biol* **13,** 616–630 (2012).

159. Hata, A. & Chen, Y.-G. TGF-β Signaling from Receptors to Smads. *Cold Spring Harb Perspect Biol* **8,** a022061 (2016).

160. Monsivais, D., Matzuk, M. M. & Pangas, S. A. The TGF-β Family in the Reproductive Tract. *Cold Spring Harb Perspect Biol* **9,** (2017).

161. Sun, P. D. & Davies, D. R. The cystine-knot growth-factor superfamily. *Annu Rev Biophys Biomol Struct* **24,** 269–291 (1995).

162. Wakefield, L. M. & Hill, C. S. Beyond TGFβ: roles of other TGFβ superfamily members in cancer. *Nat Rev Cancer* **13,** 328–341 (2013).

163. Massagué, J., Blain, S. W. & Lo, R. S. TGFbeta signaling in growth control, cancer, and heritable disorders. *Cell* **103,** 295–309 (2000).

164. de Caestecker, M. The transforming growth factor-beta superfamily of receptors. *Cytokine Growth Factor Rev.* **15,** 1–11 (2004).

165. Loomans, H. A. & Andl, C. D. Activin receptor-like kinases: a diverse family playing an important role in cancer. *Am J Cancer Res* **6,** 2431–2447 (2016).

166. Massagué, J. & Chen, Y. G. Controlling TGF-beta signaling. *Genes & Development* **14,** 627–644 (2000).

167. Robertson, I. B. *et al.* Latent TGF-β-binding proteins. *Matrix Biol.* **47,** 44–53 (2015).

168. Huang, Z., Chen, X. & Chen, D. Myostatin: a novel insight into its role in metabolism, signal pathways, and expression regulation. *Cellular Signalling* **23,** 1441–1446 (2011).

169. Zhu, X., Topouzis, S., Liang, L.-F. & Stotish, R. L. Myostatin signaling through Smad2, Smad3 and Smad4 is regulated by the inhibitory Smad7 by a negative feedback

mechanism. *Cytokine* **26,** 262–272 (2004).

170. Lee, S. J. & McPherron, A. C. Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 9306–9311 (2001).

171. Mulder, K. M. Role of Ras and Mapks in TGFbeta signaling. *Cytokine Growth Factor Rev.* **11,** 23–35 (2000).

172. Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425,** 577–584 (2003).

173. Zhang, Y. E. Non-Smad pathways in TGF-beta signaling. *Cell Res* **19,** 128–139 (2009).

174. Blobe, G. C., Schiemann, W. P. & Lodish, H. F. Role of transforming growth factor beta in human disease. *N Engl J Med* **342,** 1350–1358 (2000).

175. Pohlers, D. *et al.* TGF-beta and fibrosis in different organs - molecular pathway imprints. *Biochim. Biophys. Acta* **1792,** 746–756 (2009).

176. Verrecchia, F. & Mauviel, A. Transforming growth factor-beta signaling through the Smad pathway: role in extracellular matrix gene expression and regulation. *Journal of Investigative Dermatology* **118,** 211–215 (2002).

177. Leask, A. & Abraham, D. J. TGF-beta signaling and the fibrotic response. *FASEB J.* **18,** 816–827 (2004).

178. Meng, X.-M., Nikolic-Paterson, D. J. & Lan, H. Y. TGF-β: the master regulator of fibrosis. *Nat Rev Nephrol* **12,** 325–338 (2016).

179. Chegini, N., Luo, X., Ding, L. & Ripley, D. The expression of Smads and transforming growth factor beta receptors in leiomyoma and myometrium and the effect of gonadotropin releasing hormone analogue therapy. *Molecular and Cellular Endocrinology*

209, 9–16 (2003).

180. Wolańska, M. & Bańkowski, E. Transforming growth factor beta and platelet-derived growth factor in human myometrium and in uterine leiomyomas at various stages of tumour growth. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **130,** 238–244 (2007).

181. Lee, B. S. & Nowak, R. A. Human leiomyoma smooth muscle cells show increased expression of transforming growth factor-beta 3 (TGF beta 3) and altered responses to the antiproliferative effects of TGF beta. *J. Clin. Endocrinol. Metab.* **86,** 913–920 (2001).

182. De Falco, M. *et al.* Preoperative treatment of uterine leiomyomas: clinical findings and expression of transforming growth factor-beta3 and connective tissue growth factor. *J. Soc. Gynecol. Investig.* **13,** 297–303 (2006).

183. Arici, A. & Sozen, I. Transforming growth factor-beta3 is expressed at high levels in leiomyoma where it stimulates fibronectin expression and cell proliferation. *Fertil. Steril.* **73,** 1006–1011 (2000).

184. Joseph, D., Malik, M., Norian, J., Webb, J. & Catherino, W. Myometrial cells undergo fibrotic transformation under the influence of transforming growth factor beta 3. *Fertil. Steril.* **88,** S216–S217 (2007).

185. Norian, J. M. *et al.* Transforming growth factor beta3 regulates the versican variants in the extracellular matrix-rich uterine leiomyomas. *Reprod Sci* **16,** 1153–1164 (2009).

186. Ciebiera, M. *et al.* Ulipristal acetate decreases transforming growth factor β3 serum and tumor tissue concentrations in patients with uterine fibroids. *Fertil. Steril.* **109,** 501–507.e2 (2018).

187. Islam, M. S. *et al.* Role of activin-A and myostatin and their signaling pathway in human

myometrial and leiomyoma cell function. *J. Clin. Endocrinol. Metab.* **99,** E775–85 (2014).

188. Ciarmela, P. *et al.* Ulipristal acetate modulates the expression and functions of activin a in leiomyoma cells. *Reprod Sci* **21,** 1120–1125 (2014).

189. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Molecular Cell* **58,** 586–597 (2015).

190. Soon, W. W., Hariharan, M. & Snyder, M. P. High-throughput sequencing for biology and medicine. *Mol. Syst. Biol.* **9,** 640–640 (2013).

191. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **30,** 418–426 (2014).

192. Mehine, M. *et al.* Characterization of Uterine Leiomyomas by Whole-Genome Sequencing. *N Engl J Med* **369,** 43–53 (2013).

193. Yogev, O., Naamati, A. & Pines, O. Fumarase: a paradigm of dual targeting and dual localized functions. *FEBS J.* **278,** 4230–4242 (2011).

194. Sass, E., Blachinsky, E., Karniely, S. & Pines, O. Mitochondrial and cytosolic isoforms of yeast fumarase are derivatives of a single translation product and have identical amino termini. *J. Biol. Chem.* **276,** 46111–46117 (2001).

195. Suzuki, T., Sato, M., Yoshida, T. & Tuboi, S. Rat liver mitochondrial and cytosolic fumarases with identical amino acid sequences are encoded from a single gene. *J. Biol. Chem.* **264,** 2581–2586 (1989).

196. Jiang, Y. *et al.* Local generation of fumarate promotes DNA repair through inhibition of histone H3 demethylation. *Nat. Cell Biol.* **17,** 1158–1168 (2015).

197. Sulkowski, P. L. *et al.* Krebs-cycle-deficient hereditary cancer syndromes are defined by

defects in homologous-recombination DNA repair. *Nat. Genet.* **50,** 1086–1092 (2018).

198. Yogev, O. *et al.* Fumarase: a mitochondrial metabolic enzyme and a cytosolic/nuclear component of the DNA damage response. *PLoS Biol* **8,** e1000328 (2010).

199. O'Flaherty, L. *et al.* Dysregulation of hypoxia pathways in fumarate hydratase-deficient cells is independent of defective mitochondrial metabolism. *Hum. Mol. Genet.* **19,** 3844–3851 (2010).

200. Toro, J. R. *et al.* Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America. *Am. J. Hum. Genet.* **73,** 95–106 (2003).

201. Launonen, V. *et al.* Inherited susceptibility to uterine leiomyomas and renal cell cancer. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 3387–3392 (2001).

202. Gross, K. L. *et al.* Involvement of fumarate hydratase in nonsyndromic uterine leiomyomas: genetic linkage analysis and FISH studies. *Genes, Chromosomes and Cancer* **41,** 183–190 (2004).

203. Harrison, W. J. *et al.* Fumarate Hydratase-deficient Uterine Leiomyomas Occur in Both the Syndromic and Sporadic Settings. *Am. J. Surg. Pathol.* **40,** 599–607 (2016).

204. Tomlins, S. A. *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448,** 595–599 (2007).

205. Rowley, J. D. Chromosome translocations: dangerous liaisons revisited. *Nat Rev Cancer* **1,** 245–250 (2001).

206. Winnepenninckx, B., Rooms, L. & Kooy, R. F. Mental Retardation: A Review of the Genetic Causes. *The British Journal of Development Disabilities* **49,** 29–44 (2013).

207. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14,** 417–422 (1998).

208. Medeiros, F. *et al.* Frequency and characterization of HMGA2 and HMGA1 rearrangements in mesenchymal tumors of the lower genital tract. *Genes, Chromosomes and Cancer* **46,** 981–990 (2007).

209. Schoenmakers, E. F. *et al.* Recurrent rearrangements in the high mobility group protein gene, HMGI-C, in benign mesenchymal tumours. *Nat. Genet.* **10,** 436–444 (1995).

210. Ashar, H. R. *et al.* Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains. *Cell* **82,** 57–65 (1995).

211. Fedele, M. *et al.* Role of the high mobility group A proteins in human lipomas. *Carcinogenesis* **22,** 1583–1591 (2001).

212. Shaffer, L. G. & Lupski, J. R. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu. Rev. Genet.* **34,** 297–329 (2000).

213. Zhang, F., Carvalho, C. M. B. & Lupski, J. R. Complex human chromosomal and genomic rearrangements. *Trends Genet.* **25,** 298–307 (2009).

214. Pai, G. S., Thomas, G. H., Mahoney, W. & Migeon, B. R. Complex chromosome rearrangements. Report of a new case and literature review. *Clin. Genet.* **18,** 436–444 (1980).

215. Batista, D. A., Pai, G. S. & Stetten, G. Molecular analysis of a complex chromosomal rearrangement and a review of familial cases. *Am. J. Med. Genet.* **53,** 255–263 (1994).

216. Warburton, D. De novo balanced chromosome rearrangements and extra marker

chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am. J. Hum. Genet.* **49,** 995–1013 (1991).

217. Giardino, D. *et al.* De novo balanced chromosome rearrangements in prenatal diagnosis. *Prenat. Diagn.* **29,** 257–265 (2009).

218. Forment, J. V., Kaidi, A. & Jackson, S. P. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat Rev Cancer* **12,** 663–670 (2012).

219. Zhang, C.-Z., Leibowitz, M. L. & Pellman, D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes & Development* **27,** 2513–2530 (2013).

220. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).

221. Garcia-Torres, R., Cruz, D., Orozco, L., Heidet, L. & Gubler, M. C. Alport syndrome and diffuse leiomyomatosis. Clinical aspects, pathology, molecular biology and extracellular matrix studies. A synthesis. *Nephrologie* **21,** 9–12 (2000).

222. Heidet, L. *et al.* Somatic deletion of the 5' ends of both the COL4A5 and COL4A6 genes in a sporadic leiomyoma of the esophagus. *Am. J. Pathol.* **152,** 673–678 (1998).

223. Bustin, M. Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends in Biochemical Sciences* **26,** 152–153 (2001).

224. Reeves, R. High mobility group (HMG) proteins: Modulators of chromatin structure and DNA repair in mammalian cells. *DNA Repair (Amst.)* **36,** 122–136 (2015).

225. Reeves, R. Nuclear functions of the HMG proteins. *Biochim. Biophys. Acta* **1799,** 3–14 (2010).

226. Merika, M. & Thanos, D. Enhanceosomes. *Curr. Opin. Genet. Dev.* **11,** 205–208 (2001).

227. Lee, Y. S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes & Development* **21,** 1025–1030 (2007).

228. Thuault, S. *et al.* Transforming growth factor-beta employs HMGA2 to elicit epithelial-mesenchymal transition. *The Journal of Cell Biology* **174,** 175–183 (2006).

229. Thuault, S. *et al.* HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. *J. Biol. Chem.* **283,** 33437–33446 (2008).

230. Kools, P. F. & Van de Ven, W. J. Amplification of a rearranged form of the high-mobility group protein gene HMGIC in OsA-CI osteosarcoma cells. *Cancer Genet. Cytogenet.* **91,** 1–7 (1996).

231. Kazmierczak, B., Bullerdiek, J., Pham, K. H., Bartnitzke, S. & Wiesner, H. Intron 3 of HMGIC is the most frequent target of chromosomal aberrations in human tumors and has been conserved basically for at least 30 million years. *Cancer Genet. Cytogenet.* **103,** 175–177 (1998).

232. Markowski, D. N. *et al.* MED12 mutations in uterine fibroids--their relationship to cytogenetic subgroups. *Int. J. Cancer* **131,** 1528–1536 (2012).

233. Bertsch, E. *et al.* MED12 and HMGA2 mutations: two independent genetic events in uterine leiomyoma and leiomyosarcoma. *Mod Pathol* **27,** 1144–1153 (2014).

234. Mello, J. B. H. *et al.* MicroRNAs involved in the HMGA2 deregulation and its co-occurrence with MED12 mutation in uterine leiomyoma. *Mol. Hum. Reprod.* **24,** 556–563 (2018).

235. Galindo, L. J. *et al.* HMGA2 and MED12 alterations frequently co-occur in uterine

leiomyomas. *Gynecol. Oncol.* **150,** 562–568 (2018).

236. Klemke, M. *et al.* Overexpression of HMGA2 in uterine leiomyomas points to its general role for the pathogenesis of the disease. *Genes, Chromosomes and Cancer* **48,** 171–178 (2009).

237. Elia, L. *et al.* The knockout of miR-143 and -145 alters smooth muscle cell maintenance and vascular homeostasis in mice: correlates with human disease. *Cell Death Differ.* **16,** 1590–1598 (2009).

238. Cordes, K. R. *et al.* miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* **460,** 705–710 (2009).

239. Kornberg, R. D. Eukaryotic transcriptional control. *Trends in Cell Biology* **9,** M46–9 (1999).

240. Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34,** 77–137 (2000).

241. Egloff, S. & Murphy, S. Cracking the RNA polymerase II CTD code. *Trends in Genetics* **24,** 280–288 (2008).

242. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development* **26,** 2119–2137 (2012).

243. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16,** 129–143 (2015).

244. Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342,** 1238724–1238724 (2013).

245. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip*

Rev Dev Biol **1,** 40–51 (2012).

246. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* (2016). doi:10.1038/nature17394

247. Kang, J. J., Auble, D. T., Ranish, J. A. & Hahn, S. Analysis of the yeast transcription factor TFIIA: distinct functional regions and a polymerase II-specific role in basal and activated transcription. *Mol. Cell. Biol.* **15,** 1234–1243 (1995).

248. Papai, G. *et al.* TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* **465,** 956–960 (2010).

249. Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA* **2,** 277–298 (2011).

250. Komarnitsky, P., Cho, E. J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Development* **14,** 2452–2460 (2000).

251. Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Molecular Cell* **3,** 405–411 (1999).

252. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277,** 19639–19648 (2002).

253. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13,** 720–731 (2012).

254. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16,** 167–177 (2015).

255. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16,** 155–166 (2015).

256. Robinson, P. J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166,** 1411–1422.e16 (2016).

257. Tsai, K. L., Tomomori-Sato, C., Sato, S. & Conaway, R. C. Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex. *Cell* **157,** 1430–1444 (2014).

258. Taatjes, D. J. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences* **35,** 315–322 (2010).

259. Knuesel, M. T., Meyer, K. D., Donner, A. J., Espinosa, J. M. & Taatjes, D. J. The Human CDK8 Subcomplex Is a Histone Kinase That Requires. *Mol. Cell. Biol.* **29,** 650–661 (2009).

260. Mäkinen, N. *et al.* MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* **334,** 252–255 (2011).

261. Je, E. M., Kim, M. R., Min, K. O., Yoo, N. J. & Lee, S. H. Mutational analysis of MED12 exon 2 in uterine leiomyoma and other common tumors. *Int. J. Cancer* **131,** E1044–E1047 (2012).

262. McGuire, M. M. *et al.* Whole Exome Sequencing in a Random Sample of North American Women with Leiomyomas Identifies MED12 Mutations in Majority of Uterine Leiomyomas. *PLoS ONE* **7,** e33251 (2012).

263. Mäkinen, N. *et al.* MED12 exon 2 mutations are common in uterine leiomyomas from South African patients. *Oncotarget* **2,** 966–969 (2011).

264. Lim, W. K. *et al.* Exome sequencing identifies highly recurrent MED12 somatic mutations in breast fibroadenoma. *Nat. Genet.* **46,** 877–880 (2014).

265. Kämpjärvi, K. *et al.* Somatic MED12 mutations in uterine leiomyosarcoma and colorectal cancer. *British Journal of Cancer* **107,** 1761–1765 (2012).

266. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44,** 685–689 (2012).

267. Rump, P. *et al.* A novel mutation in MED12 causes FG syndrome (Opitz-Kaveggia syndrome). *Clin. Genet.* **79,** 183–188 (2011).

268. Graham, J. M. & Schwartz, C. E. MED12 related disorders. *Am. J. Med. Genet. A* **161A,** 2734–2740 (2013).

269. Vulto-van Silfhout, A. T. *et al.* Mutations in MED12 cause X-linked Ohdo syndrome. *Am. J. Hum. Genet.* **92,** 401–406 (2013).

270. Risheg, H. *et al.* A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat. Genet.* **39,** 451–453 (2007).

271. Zhou, H. *et al.* MED12 mutations link intellectual disability syndromes with dysregulated GLI3-dependent Sonic Hedgehog signaling. *Proceedings of the National Academy of Sciences* **109,** 19763–19768 (2012).

272. Mittal, P. *et al.* Med12 gain-of-function mutation causes leiomyomas and genomic instability. *J. Clin. Invest.* **125,** 3280–3284 (2015).

273. Turunen, M. *et al.* Uterine leiomyoma-linked MED12 mutations disrupt mediator-associated CDK activity. *CellReports* **7,** 654–660 (2014).

274. Park, M. J. *et al.* Oncogenic exon 2 mutations in Mediator subunit MED12 disrupt

allosteric activation of cyclin C-CDK8/19. *J. Biol. Chem.* **293,** 4870–4882 (2018).

275. Russo, V. E. A. V. E. A., Martienssen, R. A. & Riggs, A. D. *Epigenetic mechanisms of gene regulation*. (Plainview, N.Y. : Cold Spring Harbor Laboratory Press, 1996).

276. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17,** 487–500 (2016).

277. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351,** 720–724 (2016).

278. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33,** 245–254 (2003).

279. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14,** 204–220 (2013).

280. Wu, T. P. *et al.* DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532,** 329–333 (2016).

281. Heyn, H. & Esteller, M. An Adenine Code for DNA: A Second Life for N6-Methyladenine. *Cell* **161,** 710–713 (2015).

282. Xie, Q. *et al.* N6-methyladenine DNA Modification in Glioblastoma. *Cell* **175,** 1228–1243.e20 (2018).

283. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13,** 484–492 (2012).

284. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249,** 1288–1290 (1990).

285. Timinskas, A., Butkus, V. & Janulaitis, A. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene* **157,** 3–11 (1995).

286. Xiao, C.-L. *et al.* N6-Methyladenine DNA Modification in the Human Genome. *Molecular Cell* **71,** 306–318.e7 (2018).

287. Cheng, X. & Blumenthal, R. M. Mammalian DNA methyltransferases: a structural perspective. *Structure* **16,** 341–350 (2008).

288. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).

289. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).

290. Maekawa, R. *et al.* Genome-wide DNA methylation analysis reveals a potential mechanism for the pathogenesis and development of uterine leiomyomas. *PLoS ONE* **8,** e66632 (2013).

291. Yamagata, Y. *et al.* Aberrant DNA methylation status in human uterine leiomyoma. *Mol. Hum. Reprod.* **15,** 259–267 (2009).

292. Sato, S. *et al.* Potential mechanisms of aberrant DNA hypomethylation on the x chromosome in uterine leiomyomas. *J. Reprod. Dev.* **60,** 47–54 (2014).

293. Navarro, A. *et al.* 5-Hydroxymethylcytosine promotes proliferation of human uterine leiomyoma: a biological link to a new epigenetic modification in benign tumors. *J. Clin. Endocrinol. Metab.* **99,** E2437–45 (2014).

294. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184,**

868–871 (1974).

295. Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98,** 285–294 (1999).

296. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Letters* **589,** 2914–2922 (2015).

297. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* **13,** 436–447 (2012).

298. Yasuhara, J. C. & Wakimoto, B. T. Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* **22,** 330–338 (2006).

299. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19,** 229–244 (2018).

300. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–412 (2007).

301. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32,** 42–56 (2016).

302. Peterson, C. L. & Laniel, M.-A. Histones and histone modifications. *Current Biology* **14,** R546–51 (2004).

303. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7,** a025064 (2015).

304. Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. & Zhao, Y. SnapShot: histone modifications. *Cell* **159,** 458–458.e1 (2014).

305. Sims, R. J., Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19,** 629–639 (2003).

306. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Current Opinion in Cell Biology* **14,** 286–298 (2002).

307. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13,** 59–69 (2011).

308. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7,** 29–59 (2006).

309. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302,** 413 (2003).

310. Ogbourne, S. & Antalis, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* **331 (Pt 1),** 1–14 (1998).

311. Bushey, A. M., Dorman, E. R. & Corces, V. G. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Molecular Cell* **32,** 1–9 (2008).

312. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–1040 (2010).

313. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat. Genet.* **39,** 730–732 (2007).

314. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160,** 554–566 (2015).

315. Dickel, D. E. *et al.* Ultraconserved Enhancers Are Required for Normal Development. *Cell* **172,** 491–499.e15 (2018).

316. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167,** 633–642.e11 (2016).

317. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172,** 650–665 (2018).

318. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).

319. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130,** 77–88 (2007).

320. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell* **49,** 825–837 (2013).

321. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107,** 21931–21936 (2010).

322. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. **39,** 311–318 (2007).

323. Dorighi, K. M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell* **66,** 568–576.e4 (2017).

324. Zabidi, M. A. & Stark, A. Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet.* **32,** 801–814 (2016).

325. Dekker, J. & Misteli, T. Long-Range Chromatin Interactions. *Cold Spring Harb Perspect Biol* **7,** a019356 (2015).

326. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Molecular Cell* **49,** 773–782 (2013).

327. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50,** 1296–1303 (2018).

328. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).

329. Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49,** 1602–1612 (2017).

330. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15,** 234–246 (2014).

331. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

332. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research* **19,** 24–32 (2009).

333. Ren, G. *et al.* CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Molecular Cell* **67,** 1049–1058.e6 (2017).

334. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* (2017). doi:10.1016/j.cell.2017.11.008

335. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28,** 1057–1068 (2010).

336. Krijger, P. H. L. & de Laat, W. Regulation of disease-associated gene expression in the

3D genome. *Nat Rev Mol Cell Biol* **17,** 771–782 (2016).

337. Mechta-Grigoriou, F., Gerald, D. & Yaniv, M. The mammalian Jun proteins: redundancy and specificity. *Oncogene* **20,** 2378–2389 (2001).

338. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer* **3,** 859–868 (2003).

339. Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* **4,** E131–6 (2002).

340. Kurokawa, H. *et al.* Structural basis of alternative DNA recognition by Maf transcription factors. *Mol. Cell. Biol.* **29,** 6232–6244 (2009).

341. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *J. Cell. Sci.* **117,** 5965–5973 (2004).

342. Eychène, A., Rocques, N. & Pouponnot, C. A new MAFia in cancer. *Nat Rev Cancer* **8,** 683–693 (2008).

343. Szalóki, N., Krieger, J. W., Komáromi, I., Tóth, K. & Vámosi, G. Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling. *Mol. Cell. Biol.* **35,** 3785–3798 (2015).

344. Chinenov, Y. & Kerppola, T. K. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* **20,** 2438–2452 (2001).

345. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43,** 264–268 (2011).

346. Uht, R. M. Transcriptional Activities of Estrogen and Glucocorticoid Receptors Are

Functionally Integrated at the AP-1 Response Element. *Endocrinology* **138,** 2900–2908 (1997).

347. Pedraza-Alva, G., Zingg, J. M. & Jost, J. P. AP-1 binds to a putative cAMP response element of the MyoD1 promoter and negatively modulates MyoD1 expression in dividing myoblasts. *J. Biol. Chem.* **269,** 6978–6985 (1994).

348. Trouche, D. *et al.* Repression of c-fos promoter by MyoD on muscle cell differentiation. *Nature* **363,** 79–82 (1993).

349. Bengal, E. *et al.* Functional antagonism between c-Jun and MyoD proteins: A direct physical association. *Cell* **68,** 507–519 (1992).

350. Cao, Y. *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Developmental Cell* **18,** 662–674 (2010).

351. Zhang, Z. *et al.* The Autoregulatory Feedback Loop of MicroRNA-21/Programmed Cell Death Protein 4/Activation Protein-1 (MiR-21/PDCD4/AP-1) as a Driving Force for Hepatic Fibrosis Development. *J. Biol. Chem.* **288,** 37082–37093 (2013).

352. Wernig, G. *et al.* Unifying mechanism for different fibrotic diseases. *Proceedings of the National Academy of Sciences* **114,** 4757–4762 (2017).

353. Eferl, R. *et al.* Development of pulmonary fibrosis through a pathway involving the transcription factor Fra-2/AP-1. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 10525–10530 (2008).

354. Rajasekaran, S., Vaz, M. & Reddy, S. P. Fra-1/AP-1 transcription factor negatively regulates pulmonary fibrosis in vivo. *PLoS ONE* **7,** e41611 (2012).

355. Skubitz, K. M. & Skubitz, A. P. N. Differential gene expression in uterine leiomyoma. *J. Lab. Clin. Med.* **141,** 297–308 (2003).

356. Gustavsson, I. *et al.* Tissue differences but limited sex steroid responsiveness of c-fos and c-jun in human fibroids and myometrium. *Mol. Hum. Reprod.* **6,** 55–59 (2000).

357. Lessl, M. *et al.* Comparative messenger ribonucleic acid analysis of immediate early genes and sex steroid receptors in human leiomyoma and healthy myometrium. *J. Clin. Endocrinol. Metab.* **82,** 2596–2600 (1997).

358. Swartz, C. D., Afshari, C. A., Yu, L., Hall, K. E. & Dixon, D. Estrogen-induced changes in IGF-I, Myb family and MAP kinase pathway genes in human uterine leiomyoma and normal uterine smooth muscle cell lines. *Mol. Hum. Reprod.* **11,** 441–450 (2005).

359. Arslan, A. A. *et al.* Gene expression studies provide clues to the pathogenesis of uterine leiomyoma: new evidence and a systematic review. *Hum. Reprod.* **20,** 852–863 (2005).

360. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. **13,** 919–922 (2016).

361. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).

362. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47,** 1179–1186 (2015).

363. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47,** 598–606 (2015).

364. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer–promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49,** 1522–1528 (2017).

365. Siersbæk, R. *et al.* Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Molecular Cell* **66,** 420–435.e5 (2017).

366. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).

367. Nadine Markowski, D. *et al.* Cell cultures in uterine leiomyomas: rapid disappearance of cells carrying MED12 mutations. *Genes, Chromosomes and Cancer* **53,** 317–323 (2014).

368. Bloch, J., Holzmann, C., Koczan, D., Helmke, B. M. & Bullerdiek, J. Factors affecting the loss of MED12-mutated leiomyoma cells during in vitro growth. *Oncotarget* **8,** 34762–34772 (2017).

369. Skene, P. J. & Henikoff, S. A simple method for generating high-resolution maps of genome-wide protein binding. *Elife* **4,** e09225 (2015).

370. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* 1–10 (2017). doi:10.1016/j.ymeth.2017.04.004

371. Nagano, T. *et al.* Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc* **10,** 1986–2003 (2015).

372. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).

373. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

374. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* **42,** e168–e168 (2014).

375. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor.

Methods Mol. Biol. **1418,** 335–351 (2016).

376. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

377. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012).

378. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).

379. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

380. Vera Alvarez, R., Pongor, L. S., Mariño-Ramírez, L. & Landsman, D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35,** 1960–1962 (2019).

381. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6,** e21800 (2011).

382. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10,** 1523 (2019).

383. Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M. B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.* **33,** 722–729 (2015).

384. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Research* **22,** 2008–2017 (2012).

385. Martin, M. Cutadapt removes adapter sequences from highthroughput sequencing reads.

EMBnet.journal **17,** 10–12 (2011).

386.   Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

387.   Stark, R. & Brown, G. *DiffBind: differential binding analysis of ChIP-Seq peak data. Bioconductor. 2011.*

388.   Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44,** W160–5 (2016).

389.   Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* **4,** (2015).

390.   Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17,** 127 (2016).

391.   Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

392.   Cairns, J., Orchard, W. R., Malysheva, V. & Spivakov, M. Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz450

393.   Cao, Z., Wu, X., Yen, L., Sweeney, C. & Carraway, K. L. Neuregulin-induced ErbB3 downregulation is mediated by a protein stability cascade involving the E3 ubiquitin ligase Nrdp1. *Mol. Cell. Biol.* **27,** 2180–2188 (2007).

394.   Jens, M. & Rajewsky, N. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat Rev Genet* **16,** 113–126 (2015).

395. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18,** 1435–1440 (2011).

396. Zu, L. *et al.* The feedback loop between miR-124 and TGF-β pathway plays a significant role in non-small cell lung cancer metastasis. *Carcinogenesis* **37,** 333–343 (2016).

397. Palumbo-Zerr, K. *et al.* Orphan nuclear receptor NR4A1 regulates transforming growth factor-β signaling and fibrosis. *Nat. Med.* **21,** 150–158 (2015).

398. Kato, M. *et al.* TGF-β activates Akt kinase through a microRNA-dependent amplifying circuit targeting PTEN. *Nat. Cell Biol.* **11,** 881–889 (2009).

399. Telford, D. J. & Stewart, B. W. Micrococcal nuclease: its specificity and use for chromatin analysis. *Int. J. Biochem.* **21,** 127–137 (1989).

400. Hörz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Research* **9,** 2643–2658 (1981).

401. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research* **25,** 582–597 (2015).

402. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167,** 1369–1384.e19 (2016).

403. Blangiardo, M., Cassese, A. & Richardson, S. sdef: an R package to synthesize lists of significant features in related experiments. *BMC Bioinformatics* **11,** 270 (2010).

404. Soutourina, J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* **19,** 262–274 (2017).

405. Beekman, R. *et al.* The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24,** 868–880 (2018).

406. Xie, J. *et al.* Application of ex-vivo spheroid model system for the analysis of senescence and senolytic phenotypes in uterine leiomyoma. *Lab. Invest.* 1–13 (2018). doi:10.1038/s41374-018-0117-5

407. Xu, X. *et al.* Oxidative stress-induced miRNAs modulate AKT signaling and promote cellular senescence in uterine leiomyoma. *J. Mol. Med.* **96,** 1095–1106 (2018).

408. Vidimar, V. *et al.* The AKT/BCL-2 Axis Mediates Survival of Uterine Leiomyoma in a Novel 3D Spheroid Model. *Endocrinology* **159,** 1453–1462 (2018).

409. Tropberger, P. *et al.* Regulation of Transcription through Acetylation of H3K122 on the Lateral Surface of the Histone Octamer. *Cell* **152,** 859–872 (2013).

410. Pradeepa, M. M. *et al.* Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* **48,** 681–686 (2016).

411. Huang, D., Petrykowska, H. M., Miller, B. F., Elnitski, L. & Ovcharenko, I. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. *Genome Research* **29,** 657–667 (2019).

412. Guo, Y. *et al.* CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. *Genome Biol.* **19,** 160–17 (2018).

413. Zhao, W. *et al.* NF-κB- and AP-1-mediated DNA looping regulates osteopontin transcription in endotoxin-stimulated murine macrophages. *The Journal of Immunology* **186,** 3173–3179 (2011).

414. Phanstiel, D. H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Molecular Cell* **67,** 1037–1048.e6 (2017).

415. Grossman, S. R. *et al.* Systematic dissection of genomic features determining transcription

factor binding and enhancer function. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E1291–E1300 (2017).

416. Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Molecular Cell* **68,** 1067–1082.e12 (2017).

417. Fonseca, G. J. *et al.* Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages. *Nat Commun* **10,** 414 (2019).

418. Park, M. J. *et al.* Mediator Kinase Disruption in MED12-Mutant Uterine Fibroids From Hispanic Women of South Texas. *J. Clin. Endocrinol. Metab.* **103,** 4283–4292 (2018).

419. Poss, Z. C. *et al.* Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *CellReports* **15,** 436–450 (2016).