UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR KARDIOGENETIK

Identification of genetic risk variants for atherosclerosis using oxidative stress assays in vascular smooth muscle cells and bioinformatic approaches

Identification genetischer Risikovarianten für Artheriosklerose via oxidativem Stress Assay in glatten Muskulaturzellen und bioinformatische Ansätze

Masterarbeit

verfasst am
Institut für Kardiogenetik

im Rahmen des Studiengangs
Molecular Life Science
der Universität zu Lübeck

vorgelegt von
Torben Falk

ausgegeben und betreut von
Prof. Dr. Jeanette Erdmann

mit Unterstützung von
Dr. Tobias Reinberger

Lübeck, den 21. Juli 2022

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

_____

Torben Falk

Zusammenfassung

Ich muss das Ding wohl irgendwann auch noch in Deutsch schreiben...

Abstract

Placeholder

Acknowledgements

Daaanke an alle!

Contents

1

Introduction

1.1 Coronary artery disease

Coronary artery disease (CAD) is one of the leading causes of death in western societies, demonstrated by a prevalence in 6.7 % of American adults and leading to the annual death of 350,000 people in the USA in 2019 (Disease Control and Prevention, 2022; Fryar, 2012). CAD is characterized by the build-up of fatty plaques in the arteries supplying the heart with oxygen. This process, called atherosclerosis can interrupt the blood supply to the heart (National Health Service, 2017). Its most common complication is myocardial infarction (MI) which usually manifests as chest pain (angina) and may cause severe damage to the heart muscle. Long time, CAD can lead to heart failure (HF), the hearts' inability to pump blood properly. Next to common and well-known lifestyle factors like tobacco use or physical inactivity, CAD risk has a hereditary component (Montalescot et al., 2013).

1.2 VSMCs in CAD

A typical blood vessel is constructed from three distinct layers surrounding the lumen. The outer adventitia is rich in connective tissue and shapes the vessel. It wraps the media, the middle layer containing vascular smooth muscle cell (SMC)s (vSMCs), which are required to mediate vasodilation and vasoconstriction according to signaled requirements. The inner layer consists of endothelial cells that define the vessel's lumen. (Tucker et al., 2022; Yap et al., 2021)
(cells involved in CAD) For a long time, the role of vSMCs in the development and progression of atherosclerosis has been underestimated and over-simplified. vSMCs have been considered either to be promotive of atherosclerosis progression or beneficial for plaque stability. Only with the emergence of novel and exciting technologies like single-cell (sc) transcriptomics and lineage tracking has this model changed into a more multifaceted one. (Liu and Gomez, 2019; Grootaert and Bennett, 2021; Yap et al., 2021) The study of vSMCs in arterosclerosis is rapidly evolving, and the underlying models are being adjusted accordingly. The black and white idea of vSMCs in arterosclerosis existing either as a differentiated (contractile) phenotype or as a dedifferentiated (synthetic) phenotype is shifting towards the consideration of a diverse set of dedifferentiated phenotypes (Liu and Gomez, 2019; Grootaert and Bennett, 2021; Yap et al., 2021). The phenotypic switch describes the loss

of contractile markers and can give rise to a diverse bouquet of different phenotypes, which can be found in the fibrous cap and plaque core (Grootaert and Bennett, 2021). The characterization of these dedifferentiated phenotypes and their impact on disease progression is still the subject of intensive research.
Among others, two external stimuli that seem to play central roles as cytokines determining the fate of vSMCs in atherogenesis are transforming Growth Factor beta (TGFβ) and platelet-derived growth factor (PDGF)-BB (PDGF-BB).

1.3 TGFβ Signaling

TGFβ Signaling in General

The term transforming Growth Factor beta (TGFβ) describes a superfamily of cytokines, the most prominent of which is TGFβ1. After secretion and activation, the active TGFβ dimer binds to a heteromeric receptor complex. The intracellular signaling is mainly implemented via Smad transcription factors. The cellular effects of TGFβ are highly dependent on the cell type and can even be pleiotropic for cells of the same type. The most prominent function of TGFβ is its role in the anti-inflammatory regulation of immune cells. (Goumans and Dijke, 2018; Batlle and Massagué, 2019)

TGFβ Signaling in VSMCs & atherosclerosis

In the context of vSMCs, TGFβ promotes proliferation and hypertrophy. Further, it promotes vSMC differentiation via elevated gene expression of contractile marker genes. Additionally, TGFβ mediates the decreased expression of Kruppel-like factor 4 (KLF4) (Davis-Dusenbery et al., 2011), a transcription factor (TF) prominent for its application in inducing pluripotency (Takahashi et al., 2007). This way, TGFβ hinders (Davis-Dusenbery et al., 2011) or potentially reverses phenotype switching (Pan et al., 2020).

1.4 PDGF Signaling

PDGF Signaling in General

Five PDGF isoforms (PDGF-AA, PDGF-AB, PDGF-BB, PDGF-CC & PDGF-DD) have been identified as a dimeric combination of four distinct polypeptide chains. All five isoforms bind to tyrosine kinase receptors. Upon activation, the PDGF receptor (PDGFR) dimerizes, allowing autophosphorylation, which activates the kinase domain and creates binding sites for signaling molecules. The active receptor is involved in a plethora of prominent messaging pathways like the mitogen activated protein (MAP)-kinase pathway, phosphatidylinositol 3'-kinase (PI3K)-signaling or signal transducers and activators of transcription (STAT)-signaling. These pathways ultimately promote cellular proliferation, survival, and migration (Chen et al., 2013; Heldin, 2013; Hu and Huang, 2015).
The predominantly PDGF isoform expressed by endothelial cells has been demonstrated to be PDGF-BB (Andrae et al., 2008; Heldin, 2013) which acts as a paracrine activator for vSMCs and other mesenchymal cells (Heldin, 2013). PDGF-BB predominantly binds to PDGFRβ and plays an essential role in the development of multiple tissues, amongst others,

in the development of the cardiovascular system (Levéen et al., 1994). In adults, PDGF-BB picks up an function in wound healing processes (Robson et al., 1992). The contribution of PDGFRβ signaling in pathologic processes like cancer or cardio vascular disease has been a subject of extensive study for decades (Heldin, 2013; Raines, 2004).

PDGF Signaling in vSMCs & atherosclerosis

Similarly to the overall role of vSMCs in atherosclerosis, the role of PDGF-BB is still the subject of extensive study. In the context of vSMCs, PDGF-BB has been shown to increase KLF4 levels, which results in an increased expression of mesenchymal markers accompanied by the loss of contractile markers, serving as an external stimulus for proliferation and phenotype switching (Yap et al., 2021).
All PDGF isoforms are found in the cell wall of arteriosclerotic vessels, and the expression of PDGFR is elevated in affected vessels (Hu and Huang, 2015). For a long time, PDGF signaling and inflammation have been assumed to be promotive of disease progression (Andrae et al., 2008; Chen et al., 2013; Hu and Huang, 2015), and recently He et al. (2015) showed that PDGFRβ signaling in a mouse model leads to inflammation and increased plaque formation. In contrast to this consensus, Newman et al. (2021a) recently demonstrated that sustained signaling via PDGFRβ is required for vSMC involvement in arteriosclerotic lesions and the construction of the fibrous cap. Their mouse model shows that the lack of vSMC involvement during plaque formation can be temporarily compensated by non-vSMC-derived cells. However, long term, the lack of vSMC involvement leads to instability of arteriosclerotic lesions.

ROS in PDGF Signaling

Reactive oxygen species (ROS) are a class of highly reactive molecules derived from elemental oxygen (O_2). They are traditionally infamous for their damaging effect on proteins and nucleic acids, potentially causing irreparable damage and ultimately leading to cell death. Recently, this perception has been shifting, especially hydrogen peroxide (H_2O_2) and superoxide anion radical ($O_2^{\bullet -}$) are recognized for their role in cellular signaling. (Sies and Jones, 2020) [HOW DO THEY SIGNAL]
Human cells contain dozens of proteins, which are capable of generating ROS and enzymatically maintain a redox steady-state (Sies and Jones, 2020). H_2O_2 and $O_2^{\bullet -}$ serve as important second messengers in the central nervous system (Nayernia et al., 2014), for the repair of vascular lesions (Andrae et al., 2008), and PDGF-BB signaling (Sundaresan et al., 1995; Bouzigues et al., 2014).

1.5 GWAS

The hereditary components of disease onset and progression can provide access to its pathology on a molecular level. [MAYBE QUOTE]

GWAS

A fantastic resource for obtaining a first glance at these interactions are genome wide association studygenome wide association studies (GWAS studies), a method that enables the

identification of genetic variants associated with a phenotype.
While GWAS studies were initially an extraordinary endeavor, requiring the evaluation of hundreds or thousands of participants, they have gotten a lot more accessible with the availability of genetic data from public biobanks. The first step is profiling a suitable cohort on a genomic level and its phenotypical characterization. Subsequently, the collected data must pass through several quality control steps to remove rare variants, mismatched phenotypes, and others. Afterward, not directly analyzed variants are inferred from a reference genome. The final step of the initial analysis is implementing a statistical model; a regression model is used to test for the association of all variants with the phenotype in question. It is crucial to be completely aware of potential biases, some of which (like age, sex, or ancestry) can and need to be included as covariants in the used model. (Uffelmann et al., 2021; Flint, 2013) The model will output a list of p-values, effect sizes, and their direction for all tested variants. A GWA study is the first important step in determining causal variants for phenotype and, therefore, a glimpse into its molecular basis (Uffelmann et al., 2021).

postGWAS

Unfortunately, GWAS studies are just the first step in a long journey of establishing causal loci to gene links, uncovering the molecular basis of disease, implementing tools for clinical risk prediction, and developing treatment options. A plethora of follow-up analyses (postGWAS studies) are essential to convert the first list of exciting and associated variants into a set of credible variants and to assess the underlying molecular mechanisms.
The first necessary follow-up is fine-mapping. Due to the complex linkage disequilibrium (LD) of variants in the human genome (see section 1.6), loci identified in the GWA study usually do not contain a single variant. Instead, multiple variants in the vicinity may form a region of linked and significant variants. Fine-mapping identifies the actual causal variant(s) in this potentially complex cluster. Multiple sophisticated statistical methods have been developed. The popular approach of Bayesian modeling outputs variant-specific posterior inclusion probabilityposterior inclusion probabilities (PIPs) that form a credible set of potentially causal variant(s). Noteworthy, fine-mapping methods are continuously refined and evolve alongside the increasing complexity of the studied phenotypes. Further, fine-mapping is a statistical approach that can never determine causality. (Schaid et al., 2018; Uffelmann et al., 2021)
After identifying likely causal variants, the following steps aim to gain information on their effect in determining the analyzed phenotype. Variants require mapping to impacted genes, associated pathways, and relevant tissues and cell types, providing helpful insight into the complete picture. For these steps, no standard protocols exist, and the procedure highly depends on the genomic context of the variant of interest. Coding variants are rare but may be immediately studied on a protein level. On the other hand, non-coding variants usually greatly benefit from the consultation of more high throughput data in the form of, e.g., expression quantitative trait loci (eQTL) (Uffelmann et al., 2021).
Finally, the results and ideas derived from statistical models can and need to be taken back to the wet lab to be extended and verified. Utilizing all the recent remarkable advances in molecular and cellular biology, such as the development of increasingly comprehensive *in vitro* models and their manipulation via methods like clustered regularly interspaced short palindromic repeats (CRISPR)-Cas gene-editing (Lichou and Trynka, 2020).

1.6 Complementary High Through Put Methods

The development of high trough put methods combined with the significant increase in computational power over the last few years have paved the way for postGWA study. A short overview of some definitions and methods mentioned in this thesis can be found in the following paragraphs:

Linkage Disequilibrium

LD is a parameter used in population genetics that describes the non-random association of two or more alleles. The LD is often quantified using the correlation coefficient r^2 (Slatkin, 2008).

$$D_{AB} = p_{AB} - p_A p_B$$

$$r^2 = \frac{D_{AB}^2}{p_A(1-p_A) \times p_B(1-p_B)}$$

Where p_A and p_B are the frequency of the alleles A and B, respectively, and p_{AB} is the frequency of the AB haplotype.
The LD becomes vital in the context of GWAS studies because identified single nucleotide polymorphisms (SNPs) often do not occur in isolation but as a network of linked and significant variants that can span large haplotype blocks in the genome (Slatkin, 2008).

Locus To Gene Scores

The interpretation of GWAS data is prone to limitations and problems described in section 1.5. Link to gene (L2G) scores attempt to overcome the challenges of establishing causal relationships between variants and genes. The authors employed a machine learning model to integrate fine-mapping with functional genomics data and *in silico* predictions to link GWA study loci to their target genes. The computed L2G scores are calibrated to represent the probability $(0, 1)$. (Mountjoy et al., 2021)

Regulatory Build

The Ensembl Regulatory Build compiles a summary of putative regulatory regions found in the (human) genome. It is constructed from publically available data on epigenetic marks and TF binding sides. The build considers promotors, proximal enhancers, distal enhancers, and CCCTC binding factor (CTCF) binding sites. (Zerbino et al., 2015)

ENCODE cCRE

Similarly, the ENCyclopedia Of DNA Elements project (ENCODE) summarizes DNA accessibility and chromatin modification data into candidate cis-regulatory elements (cCREs).

Based on the existence of histone 3 lysine 4 trimethylation (H3K4me3), histone 3 lysine 27 acetylation (H3K27ac), or CTCF marks, regions showing high DNase signal are further annotated to be (Moore et al., 2020):

- proximal enhancer-like elements (pELS)
- distal enhancer-like elements (dELS)
- promoter-like elements (PLS)
- Regions with high H3K4me3 signal which might represent poised or non-canonical promotors
- CTCF-only elements

ABC Model

The activity by contact (ABC) model grants insights into potential cell-specific enhancer-gene interactions based on chromatin state, outperforming previously used methods (Fulco et al., 2019; Nasser et al., 2021).

$$ABC\ score_{E,G} = \frac{A_E \times C_{E,G}}{\sum_{all\ elements\ e\ within\ 5\ Mb\ of\ G} A_e \times C_{e,G}}$$

Generally speaking, the model incorporates the activity of an enhancer A_E and contacts with the gene of interest $C_{E,G}$, normalized by the total effect of all elements in proximity (Fulco et al., 2019; Nasser et al., 2021).

ATAC-seq

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a method to study chromatin accessibility in the genome. ATAC-seq utilizes the hyperactive Tn5 transponase to insert sequencing adapters into accessible chromatin regions. DNA is purified and amplified via polymerase chain reaction (PCR) and then sequenced. Mapping sites with insertions in the genome allows for identifying highly accessible genomic regions. (Buenrostro et al., 2013; Buenrostro et al., 2015a)
Employing PCR amplification renders ATAC-seq an extremely sensitive method. Pushing the requirement of biomaterial to the minimum, ATAC-seq is applicable on a single-cell level. For scATAC-seq, individual cells are isolated, and their DNA is tagged with barcoded primers during the PCR. These barcodes allow mapping of ATAC-seq data to the isolated cells. (Buenrostro et al., 2015b)

Hi-C & TADs

Hi-C is a method for mapping chromosomal conformation. Genome-associated proteins are cross-linked with formaldehyde, the DNA is digested with restriction enzymes, and generated overhangs are filled-in with biotinylated nucleotides. The resulting fragments are ligated, covalently linking DNA fragments initially in close spatial proximity. The DNA is purified and fragmented, allowing the pulldown of fragments containing junctions sites via the filled-in biotin tags. After sequencing the enriched fragments, their sequences are

mapped to the genome, identifying interacting DNA regions. (Lieberman-Aiden et al., 2009; Wit and Laat, 2012)
Looking at Hi-C data, TADs were identified as a fundamental feature of genome organization with an average size of 880 kb (Dixon et al., 2012; Wang et al., 2018b). What makes TADs interesting is that interactions of DNA sequences are usually confined within TADs. Tissue-specific genes and their enhancers are usually found in the middle of TADs, while the edges enrich for housekeeping genes and CTCF binding sides, which might serve as insulators between different domains (Pombo and Dillon, 2015).

1.7 Aim of the thesis

This thesis is split into two projects that both ultimately aim to contribute to a refined understanding of atherosclerosis and CAD:

- The split role of PDGF-BB during the progression of atherosclerosis (see section 1.4) indicates that PDGF-BB signaling is neither wholly beneficial nor disadvantageous to diseases, but there is an optimal stimulation dosage. Since ROS are mutually involved in PDGF-BB signaling (Sundaresan et al., 1995; Bouzigues et al., 2014) and atherosclerosis (Burtenshaw et al., 2019), we hypothesize a functional linkage. PDGF-BB signaling may cause oxidative stress, this way contributing to disease progression. The first part of this thesis will address the *in vitro* characterization of PDGF-BB stimulated vSMCs and the establishment of a robust assay for oxidative stress in vSMCs.
- High-throughput methods are a stable of modern bioscience and a great resource. The study by Aragam et al. (2021) provides a vast and important dataset for CAD research. One of the goals of this thesis is to make this data and the evaluation of its genomic context easily accessible to medical researchers in the form of an interactive web-based visualization tool. The GWA study Navigator will visualize GWA study summary statistics with different annotations in the form of gene products and associated phenotypes from other GWAS studies or putative regulatory elements.
- Finally, the data curated for the GWA study Navigator will be subject to an enrichment analysis, studying the overlap of disease-associated variants with regulatory elements in a diverse set of biosamples.

2

Material

2.1 Manufactors

Manufacturer	Seat
Agilent Technologies, Inc.	Santa Clara, CA, USA
Glaswarenfabrik Karl Hecht GmbH & Co. KG	Sondheim vor der Rhön, DE
Becton Dickinson GmbH	Heidelberg, DE
Brand GmbH & Co. KG	Wertheim, DE
Cell Applications, Inc	San Diego, CA, USA
ChemoMetec A/S	Allerod, DK
Eppendorf SE	Hamburg, DE
Heraeus Holding GmbH	Hanau, DE
Merck KGaA	Darmstadt, DE
Keyence Corporation	Osaka, JP
Kisker Biotech GmbH & Co. KG	Steinfurt, DE
Sarstedt AG & Co.	Nürnberg, DE
Sigma-Aldrich Co. LLC.	St. Louis, MO, USA
Thermo Fisher Scientific Inc.	Waltham, MA, USA
PeproTech© (Brand of Thermo Fisher Scientific Inc.)	Hamburg, DE
Pechiney Plastic Packaging, Inc.	Chicago, IL, USA
J.T.Baker® (Brand of Thermo Fisher Scientific Inc.)	Schwerte, DE
GFL mbH	Burgwedel, DE

Continued on next page

(Continued)

Manufacturer	Seat
Applied Biosystems (Brand of LUMITOS AG)	Warrington, UK
Gibco BRL (Brand of Thermo Fisher Scientific Inc.)	Gaithersburg, MD, USA
ibidi GmbH	Gräfelfing, DE
Invitrogen™AG (Brand of Thermo Fisher Scientific Inc.)	Schwerte, DE
Lonza Group AG	Basel, CHE
Bio&SELL GmbH	Feucht, DE
New England Biolabs GmbH	Ipswich, MA, USA
Nikon Corporation	Minato, JP
Sartorius AG	Göttingen, DE
SensoQuest GmbH	Göttingen, DE
Hanna Instruments Deutschland GmbH	Vöhringen, DE
Heidolph Instruments Labortechnik	Schwabach, DE
ZIEGRA Eismaschinen GmbH	Isernhagen, DE
Mettler-Toledo GmbH	Gießen, DE
Serif (Europe) Ltd.	Nottingham, UK
Microsoft Deutschland GmbH	München, DE
GitHub, Inc. (Part of Microsoft Corporation)	San Francisco, CA, USA
Python Software Foundation	Beaverton, OR, USA
JetBrains s.r.o.	Prague, CZ
The SQLite Consortium (Contact via Hipp, Wyrick & Company)	Charlotte, NC, USA

2.2 Celllines

Name	Celltype	Manufacturer
Human Aortic Smooth Muscle Cell (HAoSMC)	prim. human cell	Cell Applications, Inc

2.3 Oligonucleotides

Target	Name	Sequence
CNN1	Fw	5'-seq-3'
	Rv	5'-seq-3'
GAPDH	Fw	5'-seq-3'
	Rv	5'-seq-3'
MMP9	Fw	5'-seq-3'
	Rv	5'-seq-3'

2.4 Chemicals

Name	Manufacturer
5X First Strand Buffer	Invitrogen™AG
Antimycin A	Sigma-Aldrich Co. LLC.
BSA	Sigma-Aldrich Co. LLC.
CellROX™Green Reagent	Thermo Fisher Scientific Inc.
Collagen Type I, rat tail	ibidi GmbH
dNTP Mix	Applied Biosystems
DTT	Invitrogen™AG
Ethanol (99.9 %)	J.T.Baker®
FCCP	Sigma-Aldrich Co. LLC.
D-(+)-Glucose, 45 % solution	Sigma-Aldrich Co. LLC.
GlutaMAX™-I	Gibco BRL
Hoechst 33342,	Invitrogen™AG
IL-1β	PeproTech©
M-MLV RT (200 U/µL)	Invitrogen™AG
NAC	Sigma-Aldrich Co. LLC.
NaHCO3	Carl Roth GmbH + Co. KG
NaOH, 1 N	Carl Roth GmbH + Co. KG
Oligomycin	Sigma-Aldrich Co. LLC.
PBS	Lonza Group AG
PDGF-BB	PeproTech©
PowerUp™SYBR™GREEN Master Mix	Thermo Fisher Scientific Inc.
Roth Hexanukleotid Random-Primer	Carl Roth GmbH + Co. KG
RiboLock RNase Inhibitor (40 U/µL)	Thermo Fisher Scientific Inc.

Continued on next page

(Continued)

Name	Manufacturer
Seahorse XF calibrant	Agilent Technologies, Inc.
Sodium Pyruvate (100 mM)	Gibco BRL
TGFβ	PeproTech©

2.5 Media, Supplements

Name	Manufacturer
FBS Gold Plus	Bio&SELL GmbH
Medium 231	Gibco BRL
Smooth Muscle Cell Growth Supplement	Thermo Fisher Scientific Inc.
XF Base Medium	Agilent Technologies, Inc.

2.6 Solutions

Name	Content
IL-1β	10 µg/mL IL-1β 0.1 % BSA in PBS
NAC	0.25 M NAC in water, ~pH 7 (adjusted)
PDGF-BB	100 µg/mL PDGF-BB 0.1 % BSA in PBS
TGFβ	10 µg/mL TGFβ 0.1 % BSA in PBS

2.7 Kits

Kit	Manufacturer
Total RNA Purification Kit	Jena Bioscience GmbH

2.8 Consumables

Name	Manufacturer
384 Well Multiply PCR plates	?!?
Agilent Seahorse XF24 Cell Culture Microplate	Agilent Technologies, Inc.
Agilent Seahorse XF24 Extracellular Flux Assy Kit	Agilent Technologies, Inc.
BD Discardit™ II	Becton Dickinson GmbH
CRYSTAL qPCR-Folie	New England Biolabs GmbH
Filter tip 20 µL	Sarstedt AG & Co.
Filter tip 200 µL	Sarstedt AG & Co.
Filter tip 1000 µL	Sarstedt AG & Co.
Nunc Cell-Culture Treated Multidish 24	Thermo Fisher Scientific Inc.
Parafilm® M	Pechiney Plastic Packaging, Inc.
Pasteurpipetten ISO 7712	Glaswarenfabrik Karl Hecht GmbH & Co. KG
Pipette tip 20 µL	Sarstedt AG & Co.
Pipette tip 200 µL	Sarstedt AG & Co.
Pipette tip 1000 µL	Sarstedt AG & Co.
Quali-PCR-Tubes 0,2 mL	Kisker Biotech GmbH & Co. KG
Quali-PCR-Tubes 0,5 mL	Kisker Biotech GmbH & Co. KG
SafeSeal Gefäß 1,5 mL	Sarstedt AG & Co.
SafeSeal Gefäß 1,5 mL	Sarstedt AG & Co.
SafeSeal Gefäß 5 mL	Sarstedt AG & Co.
Serological pipette 5 mL	Sarstedt AG & Co.
Serological pipette 10 mL	Sarstedt AG & Co.
Serological pipette 25 mL	Sarstedt AG & Co.
Serological pipette 50 mL	Sarstedt AG & Co.
Spritzenfilter CHROMAFIL® PTFE, 0,20 µm	Carl Roth GmbH + Co. KG
TC Flask T75, Cell+, Vented Cap	Sarstedt AG & Co.
Tube 15 ml	Sarstedt AG & Co.
Tube 50 ml	Sarstedt AG & Co.
Via1-Casette™	ChemoMetec A/S

2.9 Devices

Name	Manufacturer
7900HT Fast Real-Time PCR System	Thermo Fisher Scientific Inc.
BZ-X810 All-in-One Flourescence Microscope	Keyence Corporation
BZ-X800 All-in-One Flourescence Microscope POWER	Keyence Corporation
Centrifuge 5415 R	Eppendorf SE
Centrifuge 5702 R	Eppendorf SE
Eclipse TS100	Nikon Corporation
Eismaschine	ZIEGRA Eismaschinen GmbH
Hera Cell	Heraeus Holding GmbH
Hera Cell 150	Heraeus Holding GmbH
Incubation/Inactivation bath 1083	GFL mbH
LA 120 S	Sartorius AG
MR 3001	Heidolph Instruments Labortechnik
NanoDrop™2000/2000c Spektralphotometer	Thermo Fisher Scientific Inc.
NucleoCounter NC-200	ChemoMetec A/S
pH 221 Mircoprocessor pH Meter	HANNA Instruments
Pipet-X	Mettler-Toledo GmbH
Reax Top	Heidolph Instruments Labortechnik
Research pipettes (2.5 µL, 10 µL, 100 µL, 1000 µL)	Eppendorf SE
Rotana 460 R	Andreas Hettich GmbH & Co. KG
SensoQuest labcycler	SensoQuest GmbH
XF24 Extracellular Flux Analyzer	Agilent Technologies, Inc.
Bench I Heraus	Heraeus Holding GmbH
Bench II Heraus	Heraeus Holding GmbH

2.10 Programs & Modules

Programs

Program	Version	Publisher
Affinity Designer	1.10	Serif (Europe) Ltd.
Excel	Version 2205	Microsoft Corporation

Continued on next page

(Continued)

Program	Version	Publisher
GitHub	-	GitHub Inc
keyence software?!		Keyence Corporation
MiKTeX	2.9	Christian Schenk
python	3.9	Python Software Foundation
PyCharm (Community edition)	2021.2.2	JetBrains s.r.o.
SchemaSpy	5.0.0	John Currier
SDS	2.2.2	Thermo Fisher Scientific GmbH
sqlite3_analyzer	3.38.5.	The SQLite Consortium
Wave Controller	2.6.3	Agilent Technologies, Inc.

Python Modules

Module	Version	Information
beautifulsoup4	4.11.1	crummy.com/software/BeautifulSoup
bokeh	2.4.1	bokeh.org
numpy	1.21.4	numpy.org
pandas	1.3.4	pandas.pydata.org
Pillow	8.4.0	python-pillow.org
pyliftover	0.4	github.com/konstantint/pyliftover
python standard library	3.9	docs.python.org
matplotlib	3.4.3	matplotlib.org
requests	2.26.0	requests.readthedocs.io
scipy	1.7.3	scipy.org
seaborn	0.11.2	seaborn.pydata.org
urllib3	1.26.7	urllib3.readthedocs.io
wget	3.2	bitbucket.org/licface/pywget

Frameworks

- This thesis was generated with the uzl-thesis class kindly provided by Prof. Till Tantau.
- Styling of the GWAS Visualizer was done with the cascading style sheets (CSS) Framework Bootstrap.

3

Methods

3.1 Cultivation and differentiation of HAoSMCs

Human aortic SMCs (HAoSMCs) were used for the following experiments.. This cell type is commonly used to study of cardiovascular function and disease ([**Reference for this claim**]). Cells were kept at 37 °C and 5 % carbon dioxide (CO_2) whenever possible. For differentiation, cells were first treated with TGFβ to induce a contractile phenotype. Afterward, HAoSMCs were stimulated with interleukin 1 beta (IL-1β) and PDGF-BB to induce a synthetic or pro-inflammatory phenotype. Please refer to sections 1.3 and 1.4 and the referenced literature for more information.

Thawing and Cultivation

For longtime storage, cells were stored in liquid nitrogen. New cells (6^{th} passage) were thawed at 37 °C in the water bath and transferred to a 15 mL tube when required. After centrifugation for two min at 300 xg, the supernatant was discarded. The cell pellet was taken up in 14 mL of Human Vascular Smooth Muscle Cell Basal Medium (Medium 231) (M231) + SMGS for cultivation in a TC Flask T75. Every other day, 2/3 of the medium was removed and replaced by fresh. Cells were cultivated to a maximum passage of 10.

Passaging

HAoSMCs were passaged at a maximum of 80 % confluency (approximately once a week). The medium was removed, and cells were washed once with 5 mL of PBS. Washed cells were incubated with 3 mL trypsin for 4 min at 37 °C. The trypsin was inactivated by adding 7 mL M231. Subsequently, the cell suspension was transferred to a 15 mL tube and pelleted for 4 min at 300xg. Finally, the supernatant was removed and the pellet resuspended in M231 supplemented with SMGS. The cells were seeded with a density of 500×10^3 HAoSMCs per TC Flask T75.

Preparation of Collagen I matrix

The collagen type I (col I) matrix (1.8 mg/mL) was prepared by mixing the components according to table 3.1. All components were stored at 4 °C, and all pipetting steps were carried out on ice:

Table 3.1: col I Matrix Composition

component	concentration	volume (µL)
H_20	-	38.9
M231	-	53.3
SMGS	20x	5,3
NaOH	1 M	2,7
NaHCO3	7.5 %	2.1
col I	5 mg/mL	57.6
total	-	160

160 µL of matrix mix were transferred in each well of a 24-well multidish. For polymerization, the matrix was incubated at 37 °C for 60 min.

Differentitation of HAoSMCs

Differentiation was carried out over seven days in 24-well multidishes. Differentiation was carried out in 1 mL M231 supplemented with 1 % FBS, and the indicated cytokines:

- **Day 0:** Matrix and cells were prepared as described in the previous section. Seeding of 40×10^3 in M231 + SMGS on 160 µL col I matrix or the Nunclon™ Delta well.
- **Day 1:** The medium was replaced with 1 mL M231 + 1 % FBS + 5 ng/mL TGFβ
- **Day 5:** The medium was replaced with 1 mL M231 + 1 % FBS + 10 ng/mL IL-1β + 10 ng/mL PDGF-BB
- **Day 7:** Potential additional assay-specific stimulation.

Controls were run alongside without cytokines.

3.2 mRNA Quantification

SYBR® Green is an intercalating DNA dye that allows for the monitoring of DNA amplification. Fluorescence is measured after every amplification cycle of the PCR. A lower quantification cycle (Cq) corresponds to a higher initial DNA concentration. (Huggett and Bustin, 2011)
Quantitative PCR (qPCR) was utilized to assess the messenger-RNA (mRNA) concentration of the two reporter genes calponin 1 (CNN1) and matrix metallopeptidase (MMP) 9 (MMP9) in differentiated HAoSMCs. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a reference.

RNA Isolation

RNA was isolated using the Total RNA Purification Kit. The extraction was performed according to the manufacturer's protocol, using the optional washing step with 700 µL 80 % ethanol, and RNA was eluted with 30 µL of RNase-free water.

Reverse Transcription

For RT, RNA samples were diluted to yield 10 µL of 10 ng/µL. The RNA was heated for 5 min at 68 °C. Afterward, 10 µL of the RT reaction mix was added:

Table 3.2: Reaction Mix for RT

component	concentration	volume (µL)
First Strand Buffer	5x	4
DTT	?!?	2
dNTP	?!?	1
Oligos	?!?	1
RiboLock	?!?	1
M-MLVRT	?!?	1

RT was carried out for 60 minutes at 37 °C before inactivating the enzyme for 5 minutes at 95 °C. Complementary DNA (cDNA) was either used for qPCR or stored at -20 °C.

qPCR

Table 3.3: Sample Composition for qPCR

component	concentration	volume (µL)
SYBR GREEN Master Mix	1:2	3.75
Primer (forward + reverse)	5 pM (each)	1.125
H_20	-	1.125
cDNA	-	1.5

Samples were prepared in a 384-well Multiply PCR plate. Wells were sealed and thoroughly mixed by invertation of the plate. The assay was performed with a 7900HT Fast Real-Time PCR System:

Table 3.4: qPCR Cycle

step	time (s)	temperature (°C)	loop to	passes
1	120	50		1
2	600	95		1
3	15	60		40
4	60	60	3	40
5	600	95		1
6	-	16		1

Processing of Data

The Cq was automatically calculated by the software SDS2.2.2. The arithmetic mean of three technical replicates was calculated for each sample, disregarding values that are apparent outliers. For normalization, the mean Cq of the reference gene GAPDH was subtracted from the mean Cq of the gene of interest:

$$\Delta ct = ct(\text{geneofinterest}) - ct(\text{GAPDH})$$

Taking into account the exponential amplification of DNA in PCR, the Δct can then be transformed into a relative expression level. Where 10×10^6 is a constant to yield readable values:

$$\text{rel.expr.} = 2^{-\Delta ct 10 \times 10^6}$$

In total, four biological replicates were performed. Data visualization and statistical analysis were carried out in python. A student's t-test was used for statistical testing, and a p-value of 0.05 is considered significant. For detailed information, please refer to the script.

3.3 Energy Profiling

The Seahorse XF Analyzer allows real-time measurement of dissolved oxygen and protons in a confined small volume using solid-state sensor probes. The measurements are then used to calculate a cell monolayer's oxygen consumption rate (OCR) and extracellular acidification rate (ECAR). The OCR and ECAR are indicators for mitochondrial respiration and glycolysis. They can be used to assess the metabolic state of cells. (AgilentTechnologies, 2022)
Seahorse assay was utilized to assess the energy profile of differentiated HAoSMCs. For this assay, cells were not differentiated in a 24-well multidish but XF24 Cell Culture Microplates. Since the confined volume required for the assay would not fit the matrix, cells were not cultivated on col I!

Seahorse Assay

On the day before the assay, the Seahorse XF Analyzer was set up to calibrate. The XF24 Extracellular Flux Assay Kit cartridge was equilibrated in Seahorse XF calibrant overnight at 37 °C in a non-CO_2 environment.
On the day of the assay, HAoSMCs were washed with 500 µL PBS each. Afterward, the XF BASE medium was supplemented with 1 mM sodium pyruvate, 10 mM glucose, 2 mM glutamine, and 90 µM NaOH. Cells were incubated for 1 h at 37 °C in a non-CO_2 environment with supplemented 500 µL XF BASE medium. Toxins for disruption of the respiratory chain were prepared and loaded into the XF24 Extracellular Flux Assay cartridge:

Table 3.5: Toxin Concentrations for XF24 Extracellular Flux Assay

component	concentration in cartridge (µM)	volume in cartridge (µL)	concentration in well (µM)
Oligomycin	14	55	1.4
FCCP	10	60	2.0
Antimycin	50	65	5.0

The cartridge was mounted into the XF Analyzer for calibration. After successful calibration, the hydration cartridge was replaced with the cell plate. The measurement was programmed as the following:

- Calibration of the probes.
- Equilibration
- 3 Repeats of:
  - Mixing (1 min)
  - Pause (2 min)
  - Detection of OCR and ECAR (4 min)
- Pause (2 min)
- Injection of 55 µL Oligomycin
- 3 Repeats of:
  - Mixing (1 min)
  - Pause (2 min)
  - Detection of OCR and ECAR (4 min)
- Pause (2 min)
- Injection of 60 µL FCCP
- 3 Repeats of:
  - Mixing (1 min)
  - Pause (2 min)
  - Detection of OCR and ECAR (4 min)
- Pause (2 min)
- Injection of 55 µL Antimycin
- 3 Repeats of:
  - Mixing (1 min)
  - Pause (2 min)
  - Detection of OCR and ECAR (4 min)

After the measurement, the medium was removed, and HAoSMCs were stained for 15 min with 1 µg/mL Hoechst 33342 in PBS. Finally, the cells were imaged to determine cell count for normalization.

Processing of Data

Cells were quantified using a python script provided by Dr. Tobias Reinberger. OCR and ECAR were calculated with Wave Controller, which were normalized using the cell count and the signal in the control wells. In total, three biological repeats were recorded. One repeat was excluded because no changes in OCR and ECAR could be detected and cells detached from the bottom of the wells. One of the five technical repeats for each condition was manually excluded for the remaining replicates. Further, initial OCR and ECAR, as well as the characteristics of the respiratory chain displayed in figure 3.6 B, were calculated using a modified python script provided by Dr. Tobias Reinberger. Welch's t-test was used for statistical analysis, and a p-value of 0.05 is considered significant. For detailed information, please refer to the script.

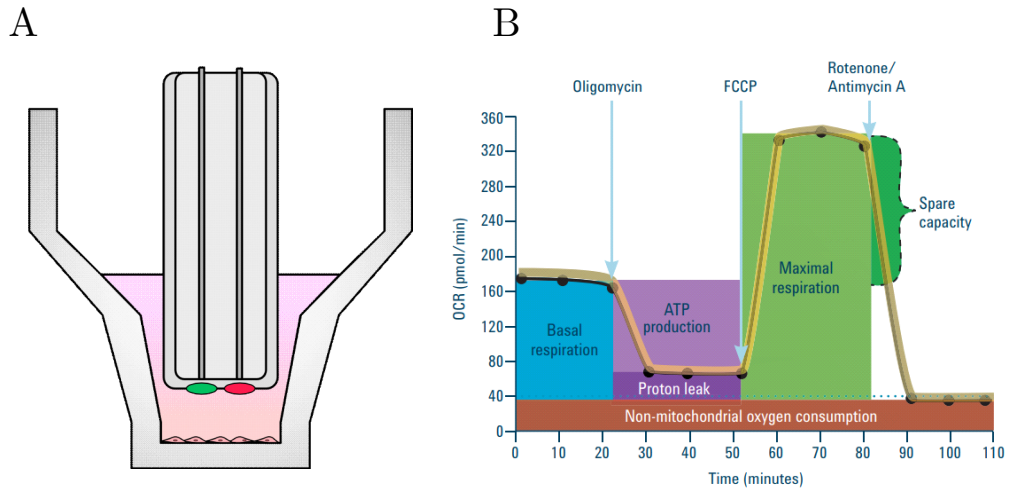

Figure 3.6: Basics of Seahorse Assay
(A) Schematic of a well used in Seahorse Assay. For the measurement, the piston in the middle lowers to the bottom, defining a restricted space at the bottom. OCR and ECAR in this volume are measured via two probes (red and green). **(B)** Exemplary curve for OCR recorded over time and extractable properties of the respiratory chain (**AGILENT WEBSITE; ADD SOURCE**).

3.4 Oxidative Stress Assay

CellROX™ Green is a fluorescent dye that gets oxidized by elevated levels of ROS. The oxidized form binds to DNA and shows bright-green fluorescence (Thermo Fisher Scientific Inc., 2022).
CellROX™ Green assay was used to assess the generation of ROS in differentiated HAoSMCs. After differentiation, further stimulation (from here on referred to as *boost*) with PDGF-BB was carried out. Finally, a recovery experiment was performed using NAC. NAC is a potent antioxidant.

CellROX™ Assay

For the assay, cells were washed with PBS. Then the boost was performed using variable concentrations of PDGF-BB in 300 µL Hanks balanced salt solution (HBSS). For ROS quenching with NAC, 0.25 M NAC solution was added to the wells 2 h prior to the experiment and the PDGF-BB-CellROX™ Green Mix.

During the boost, cells were kept at 37 °C in a 5 % CO_2 environment. The boost time (60, 120, 180, or 240 min) is indicated along the results of the respective experiment. Imaging was done with the BZ-X810 All-in-One Fluorescence Microscope, using standard sensitivity. Images for the N-acetylcysteine (NAC) quench were recorded as a z-stack and merged into one image using [KEYENCE SOFTWARE].

Table 3.7: Composition of the PDGF-BB-CellROX™ Green Mix

component	concentration	final concentration	volume (µL)
HBSS	-	-	300
PDGF	100 µg/mL	(0 - 400 ng/mL)	variable
Hoechst	1 mg/mL	1 µg/mL	0.3
CellROX™ Green (1:500)	2.5 mM	5 µM	0.6
NAC	0.25 M	(0 - 8 mM)	variable
total	-	-	~300

Processing of Data

Seven biological repeats were performed for PDGF-BB-boost titration. One repeat was excluded because of a high signal in the negative control. Four biological repeats were performed for the NAC quench, one of which has been excluded because no signal in the positive control could be detected. For quantification of signal intensity, pixels with a green value higher than 90 were counted. Differences in cell count were adjusted by dividing the green pixel by the number of blue pixels with a threshold of 80. To adjust for the large variance in total signal intensity between biological repeats, values were normalized by division through the total signal of all recorded conditions. The Mann-Whitney U test was used for statistical testing. A p-value of 0.05 is considered significant. For detailed information, please refer to the scripts.

3.5 Curation of Data for postGWAS Analyses

Data for postGWA study analyses and co-visualization with the GWA study summary statistics were downloaded from public resources. Processing of the data and further annotation is briefly listed and described below. The generated tables are summarized in figure 4.9 and table 4.8. For a complete view, please refer to the download scripts.

- **GWAS Summary Statistics:** The CAD GWA study summary statistics from Aragam et al. (2021) and a list of identified proxy SNPs were annotated via the Ensembl representational state transfer (REST) application programming interface (API) by Dr. Tobias Reinberger.
- **HGNC Gene List** The newest quarterly update to the complete Human Gene Nomenclature Consortium (HGNC) dataset was downloaded via the HGNC file transfer protocol (FTP) server. The dataset was generated a list of all 43,135 approved symbols, mapping to their HGNC ID. Further, a list of all 98,723 symbols (approved, alias, and previous) mapping to their HGNC ID was generated.
- **Linked SNPs** LD r^2 values in a 500 kb window were calculated for the list of CAD GWA study proxy variants via the Ensembl REST API. For humans, Ensembl calculates the LD with data from the 1000 Genomes project (see table 3.8). In the same process, linked SNPs were annotated with their most severe consequence from the Ensembl variant effect predictor (VEP). In total, information for 449,770 relationships were downloaded.
- **Ensembl Genome Annotations** The newest Ensembl build (Ensembl release 106) was downloaded via the Ensembl FTP server. Features annotated as protein-coding

Table 3.8: 1000 Genomes Populations

Name	Size (individuals)	Description
1000GENOMES:phase3:ALL	2504	All phase 3 individuals
1000GENOMES:phase3:AMR	347	Americans
1000GENOMES:phase3:EAS	504	East Asians
1000GENOMES:phase3:EUR	503	European
1000GENOMES:phase3:SAS	489	South Asian

(19,994), long non-coding RNA (lncRNA) (17,734), or micro RNA (miRNA) (1,877) genes were extracted. Further, gene symbols were mapped to their HGNC ID if possible.

- **Ensembl Regulatory Build** The newest Ensembl regulatory build (Ensembl release 106) was downloaded via the Ensembl FTP server. The build contains:
  - 110,623 open chromatin regions
  - 30,873 TF binding sites
  - 175,885 CTCF binding sites
  - 127,935 enhancers
  - 36,597 promotors
  - 140,548 promotor flanking regions
- **Open Target Genetics l2g Scores** The latest list of Open Target Genetics L2G Scores was downloaded via the open target genetics FTP server. Entries were annotated with their HGNC ID whenever possible. 655 entries that do not map to a gene approved by the HGNC were excluded, yielding a total of 3,580,206 database entries.
- **TSS** 35,160 transcription start sites (TSS) for protein-coding genes were extracted from a University of California Santa Cruz (USCS) Genome Browser dump.
- **Associated traits from the GWAS catalog** The SNP trait associations from the latest release of the GWAS catalog were downloaded via the GWAS catalog FTP server. 14,892 SNP-trait correlations missing a position on the human reference genome or the p-value for the association were excluded from the data set. In total, 370,002 associations from 5,831 distinct studies were collected.
- **TADs** TADs predicted by software adapted from Dixon et al. (2012) were downloaded via the 3D genome browser. In total, TADs in 40 distinct biosamples were downloaded.
- **scATAC-seq from Newman et al. (2021b)** Processed scATAC-seq data for 8 cell types [SOME MORE INFO] were scraped from the Miller Lab GitHub repository.
- **scATAC-seq from CATlas** Processed scATAC-seq data was scraped from the Ren Labs website for 222 biosamples.
- **ABC model** The ABC model data for 131 biosamples was downloaded from the Engreitz Lab FTP server. The data was further translated from Genome Reference Consortium Human Build 37 (hg19) to Genome Reference Consortium Human Build 38 (hg38) using pyliftover.
- **ENCODE cCREs** cCREs in distinct biosamples were downloaded by Dr. Tobias Reinberger, filtering out elements annotated as *unclassified*.

3.6 Visualization of GWAS data

For visualization of the data, a bokeh application was built. The application fetches the data from the database and renders it to a web browser (Bokeh Development Team, 2022). Bokeh is a python module that allows easy and interactive visualization of data. It combines the powerful data processing tools of python with the interactivity of javascript (JS) running in the browser. The python side of bokeh creates python objects that are serialized into JavaScript Object Notation (JSON) data and handed over to bokehJS. bokehJS deserializes these into JS objects that are rendered to the browser. The integrated bokeh server additionally allows synchronization of data between the underlying python environment and browser-side JS library. All in all, allowing real-time updates to the displayed data.

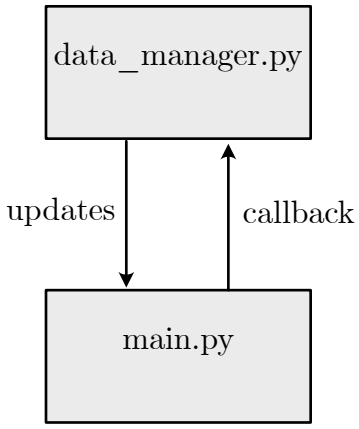

Figure 3.9: Architecture of the GWAS Navigator

According to good design principles, the concerns of the application are split into two sections, as shown in fig. 3.9. Reading the data from the database and further processing steps are managed by a data provider and enclosed in one class. In contrast to the model-controller-view architecture, a popular architectural pattern for the design of user interfaces, there is no partition between a view and a controller (Langtangen and Johansen, 2015). Since data visualization and the control widgets are created by bokeh, it is convenient to use the built-in event listeners of the library to handle the required callbacks. Therefore, the main file is responsible for creating all plots and widgets and handling user inputs.

3.7 Enrichment analysis

Based on the data in the database, initial postGWAS studies were run. Annotation enrichment analyses are a popular tool for identifying terms over-represented in a list of interest. The most prominent application is their application as gene set enrichment analysis (GES analysis). GES analyses are used to check for the overrepresentation of a candidate gene list in a predefined set of genes (Tipney and Hunter, 2010). In this case, using Fisher's exact test, the method is used to determine whether cCREs overlap with CAD-associated SNPs is enriched in certain biosamples.
For the analysis, cCREs annotated as unclassified were excluded. As a list of CAD-

associated SNPs, the list of 241 proxy variants from the database was used, as well as all linked variants ($r^2 \geq 0.6$) in the 1000 Genomes European Population. The following parameters were calculated for all biosamples:

- The number of distinct cCREs among all biosamples (m)
- The number of distinct cCREs that are annotated in the biosample of interest (mt)
- The Number of distinct cCREs that overlap with an SNP in the SNP list in any biosample (n)
- The Number of distinct cCREs that overlap with an SNP in the SNP list in the biosample of interest (nt)

The p-value for the number of overlaps greater than or equal to the observation can be calculated as the cumulative distribution function of the hypergeometric distribution ().

$$P(\sigma_t \geq n_t) = \sum_{k=n_t}^{min(m_t,n)} \frac{\binom{n}{k}\binom{m-n}{m_t-k}}{\binom{m}{m_t}}$$

To account for the multiple comparisons problem, p-values were adjusted with Bonferroni correction, where n is the number of tests (≡ number of biosamples):

$$p_{ajd.} = p * n$$

The analysis and visualization were done in python. An adjusted p-value of 0.05 is considered significant. Finally, the identified biosamples were annotated via the cell line database Cellosaurus. For detailed information, please refer to the analysis scripts.

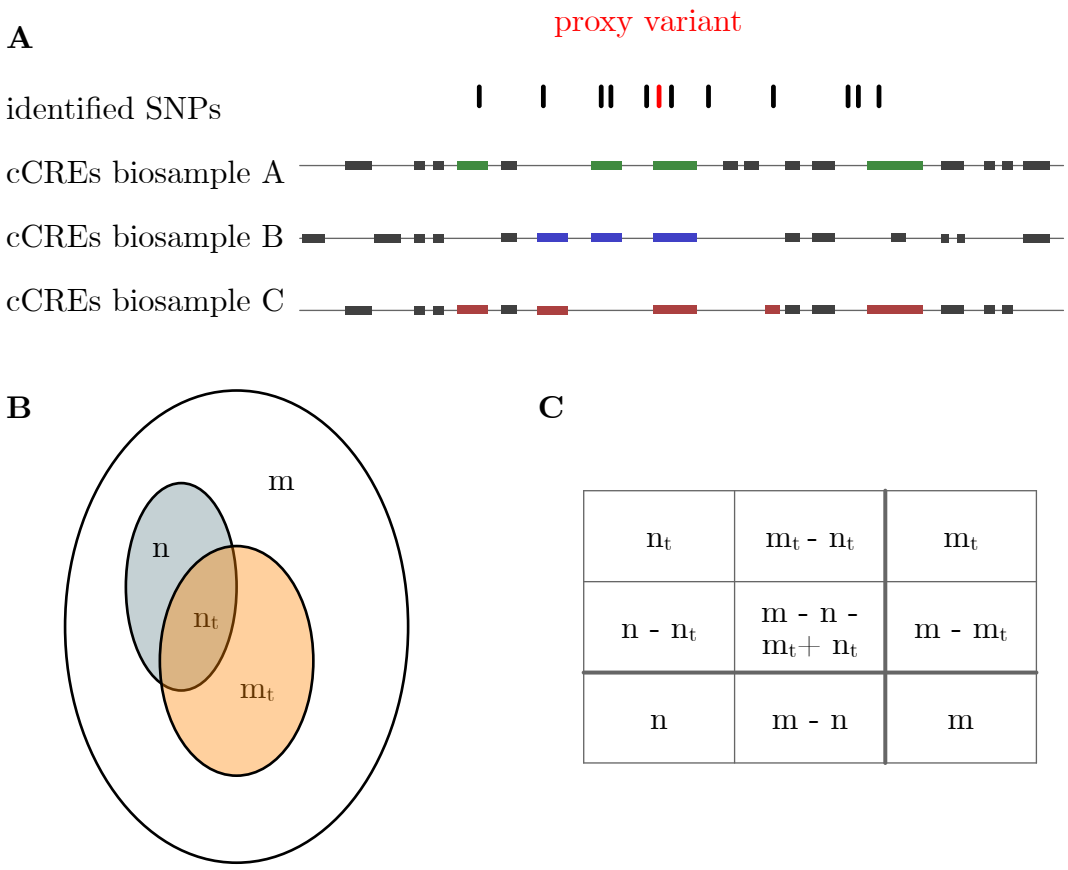

Figure 3.10: Enrichment analysis for cCREs overlapping with CAD risk SNPs

(A) Visual representation of the overlap calculation for enrichment calculation. The proxy variant is indicated as a red line, variants in LD are indicated as black lines. cCRE are shown as boxes, those that are overlapping with an SNP were colored according to the biosample they were annotated in. **(B)** Venn diagram of these values for a biosample. **(C)** Schematic contingency table for a biosample. (m) is the number of distinct cCREs found among all biosamples (23 in this example); (mt) the number of distinct cCREs annotated in the biosample of interest (16 for biosample A, 14 for biosample b, 14 for biosample C); (n) the number of distinct cCREs overlapping with an SNP (6 in this example); the number of distinct cCREs overlapping with an SNP in the biosample of interest (4 for biosample A (green), 3 for biosample B (blue), 5 for biosample C (red))

4

Results

4.1 Differentitaion

To characterize the influence of PDGF-BB stimulation on HAoSMCs, the cells were first treated with TGFβ for four days to push them towards a phenotype that resembles the contractile phenotype. Cells were stimulated for two days from this standardized starting point with IL-1β and PDGF-BB. The induced phenotypes were then characterized via qPCR and Seahorse Assay.

Expression of CNN1 and MMP9

To confirm that the HAoSMCs first adopt a contractile phenotype and to track further differentiation after stimulation with PDGF-BB, the mRNA levels of the marker genes CNN1 and MMP9 were determined using qPCR. CNN1 serves as a contractile marker and MMP9 as a marker for a synthetic phenotype. For better comparability, mRNA were normalized with expression of the housekeeping gene GAPDH.
Figure 4.1 (top panel) illustrates, stimulation of HAoSMCs cultivated on a col I-matrix with TGFβ causes a significant increase in CNN1 expression (+– vs. – –). After further stimulation with PDGF-BB and IL-1β, while not significant, CNN1 expression declines again (+– vs. ++) but is still significantly higher than in HAoSMCs which were not stimulated (– – vs. ++). A similar trend is noticable for HAoSMCs cultivated on plastic. However, this effect did not reach significance after four biolgical repeats. Additionally, stimulation of HAoSMCs on plastic with TGFβ followed by stimulation with PDGF-BB and IL-1β, yields a significantly lower expression of CNN1 (++ Matrix vs. ++ Plastic).
As seen in the bottom panel of figure 4.1, after 4 biological repeats, no statistically significant trends can be observed for the expression of MMP9. Still the average expression of MMP9 seems to increase on col I matrix compared to plastic for all conditions. The most prominent difference being between HAoSMCs treated first with TGFβ as well as with PDGF-BB and IL-1β (++ Matrix vs. ++ Plastic, $p = 0.063$).

Energy profile

In addition to the expression of CNN1 and MMP9, the energy profiles of HAoSMCs were assessed via Seahorse Assay. It is important to note that the assay was carried out on plastic

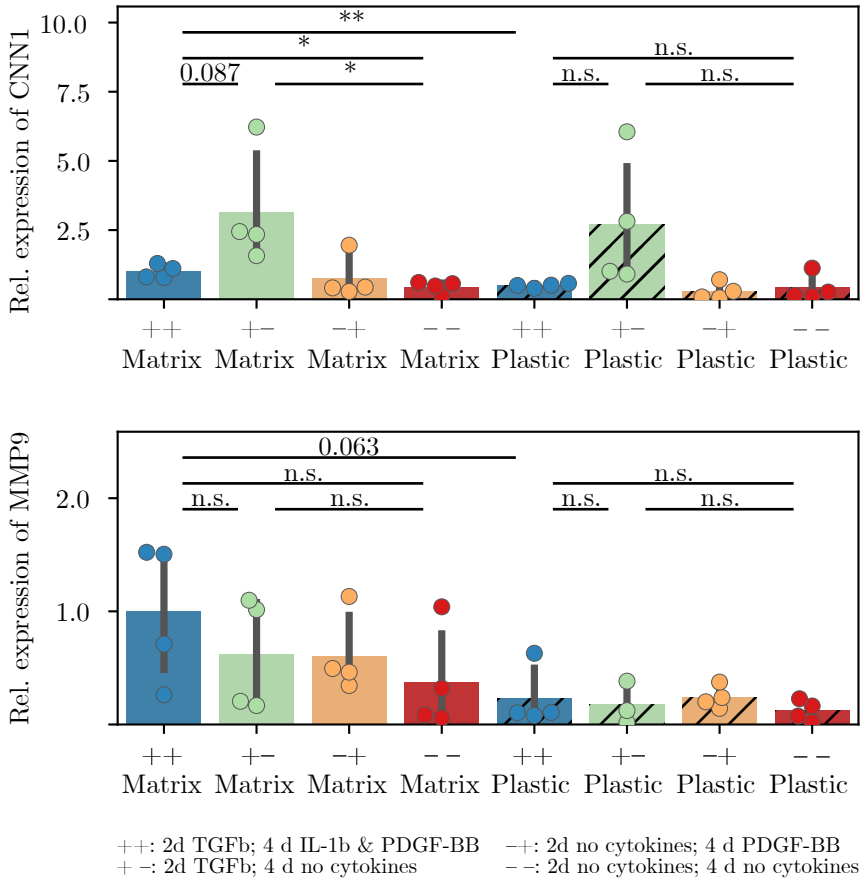

Figure 4.1: Relative Expression of CNN1 and MMP9 in HAoSMCs
qPCR analysis of expression for contractile marker CNN1 (top) and synthetic marker MMP9 (bottom) for HAoSMCs differentiated with different combinations of cytokines: **++:** four days with TGFβ followed by two days with IL-1β and PDGF-BB; **+−:** four days with TGFβ followed by two days without stimulation; **−+:** four days without stimulation followed by two days with IL-1β and PDGF-BB; **—:** six days without stimulation. All four conditions were tested on two different surfaces (plastic vs. col I matrix). Expression levels are in relation to expression of housekeeping gene GAPDH. Statistical analysis for (n = 4) biological repeats was performed using student's T-test: $* : p < 0.05; ** : p < 0.01$

because the col I matrix does not fit into the confined compartment created by the piston for detection of OCR and ECAR. Furthermore, only two biological repeats were evaluated because it became clear that all other experiments would be carried out on a col I matrix. Therefore all the following considerations should take these decisions and limitations into account.

The readout parameters of the Seahorse assay are the OCR as a representation of mitochondrial activity and the ECAR, representing the glycolytic activity of the cells. OCR and ECAR for HAoSMCs are displayed in figure 4.2. All cells show characteristic changes in OCR after addition of toxins impacting the respiratory chain (compare to figure 3.6

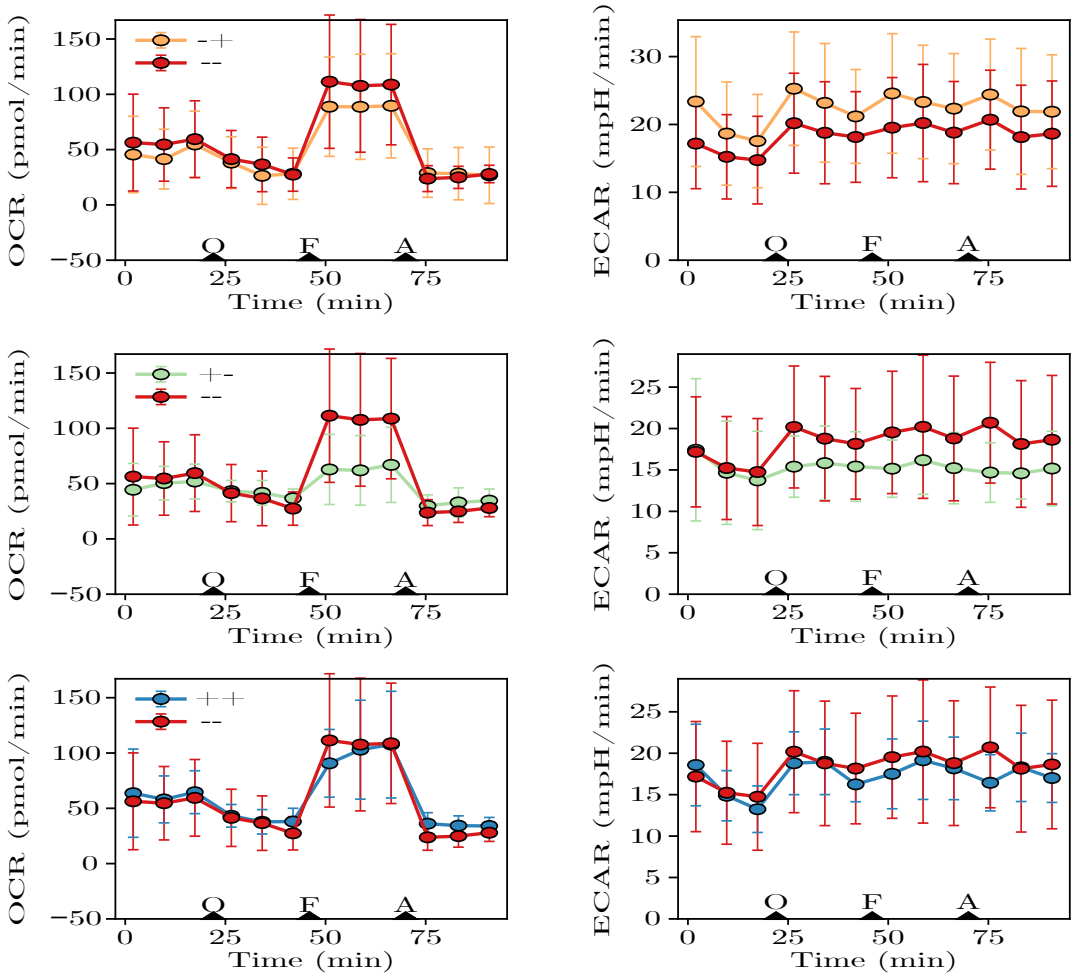

Figure 4.2: OCR and ECAR of HAoSMCs
Seahorse assay for HAoSMCs differentiated with different combinations of cytokines. **++:** four days with TGFβ followed by two days with IL-1β and PDGF-BB; **+−:** four days with TGFβ followed by two days without stimulation; **−+:** four days without stimulation followed by two days with IL-1β and PDGF-BB; **——:** six days without stimulation. OCR and ECAR are shown for −+ (top), +− (middle) and ++ (bottom) in comparison to ——. Injection times for toxins (O: Oligomycin; F: FCCP; A: Antimycin A) are marked as triangles. All tracks were recorded for cells cultivated on plastic. Shown datapoints are the average of (n = 2) biological repeats.

B). After inhibition of the ATP synthase with Oligomycin, the basal OCR drops, this way revealing the proportion of the OCR required for adenosine triphosphate (ATP) production. Subsequently, the addition of FCCP decouples the respiratory chain, destroying the proton gradient over the mitochondrial membrane. As a result, the cells reach their maximal respiratory capacity. Finally, the inhibition of coenzyme Q-cytochrome c reductase (complex III) with Antimycin A stops all mitochondrial respiratory activity.
The ECAR shows a mild increase for all conditions after the addition of Oligionmycin, most

likely because the cells are compensating for the loss of mitochondrial ATP production via increased glycolysis.

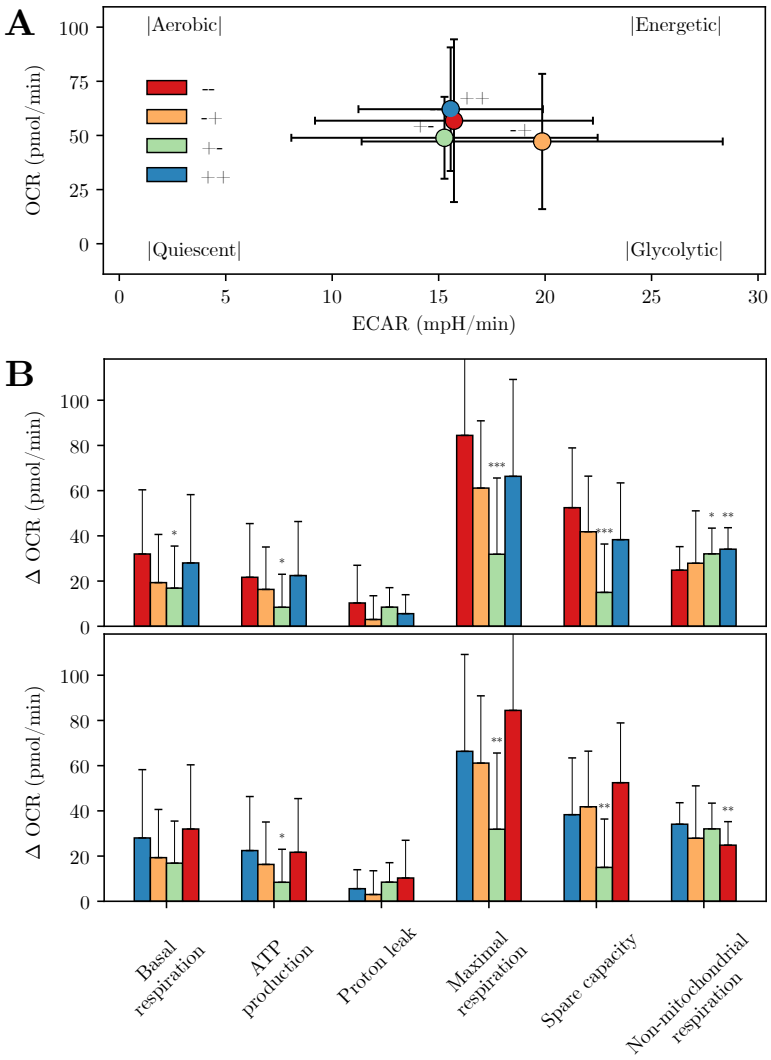

Figure 4.3: Energy profile of HAoSMCs
Seahorse assay for HAoSMCs differentiated with different combinations of cytokines as described in figure 4.2. (**A**) Initial OCR and ECAR of the four tested conditions. (**B**) Characteristics of the respiratory chain calculated from the tracks shown in figure 4.2 as described in section 3.3. Statistical analysis for (n = 2) biological repeats was performed using student's T-test: $* : p < 0.05; ** : p < 0.01, ; *** : p < 0.001$

Looking at the energy profile which describes the basal state of the differentiated cells, OCR and ECAR are quite similar for the conditions ++, +− and —. The only outlier showing a higher ECAR, are HAoSMCs only stimulated with only IL-1β and PDGF-BB (−+) (fig. 4.3, A). More interesting differences can be observed when examining characteristics of the respiratory chain. Stimulation with only TGFβ causes a significant decrease in basal respiration, ATP production, maximal respiration as well as spare capacity (figure 4.3, B

top). Further stimulation with IL-1β and PDGF-BB then causes a significant increase of these parameters to similar levels as in initially dedifferentiated HAoSMCs (figure 4.3, B bottom).

4.2 Evaluation of oxidative Stress

Finally, it was evaluated if further stimulation with PDGF-BB would stimulate the cells to generate ROS to the extent that can not be compensated by the ROS defense and lead to oxidative stress.

PDGF boost of out cells induces oxidative stress

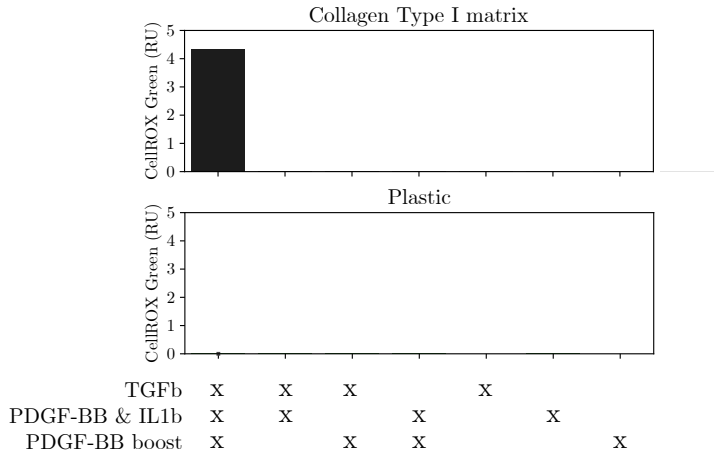

Figure 4.4: Boost with PDGF-BB induces generation of ROS.
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGFβ; followed by two days with IL-1β and PDGF-BB; followed by 2 h boost with 200 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix (top) or plastic (bottom). The hown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by Hoechst 33342 signal. No statistical analysis for (n = 1) biological repeats was performed.

At first, an experiment already done in the group was validated. Stimulating the four tested combinations for 2 additional hours with 200 ng/mL PDGF-BB in HBSS. As displayed in figure 4.4 only stimulation for four days with TGFβ, followed by two days with IL-1β and PDGF-BB, followed by a 2 h boost with PDGF-BB, was able to trigger noticeable generation of ROS for cells cultivated on col I-matrix. No generation of ROS was detectable for HAoSMCs cultivated without the col I-matrix.

Characterization of the CellROX™ Assay

To get a better understanding of the assay and its limits, a titration was carried out. For this, HAoSMCs stimulated for four days with 5 ng/mL TGFβ as well as two days with 10 ng/mL IL-1β and 10 ng/mL PDGF-BB, were boosted with different concentrations

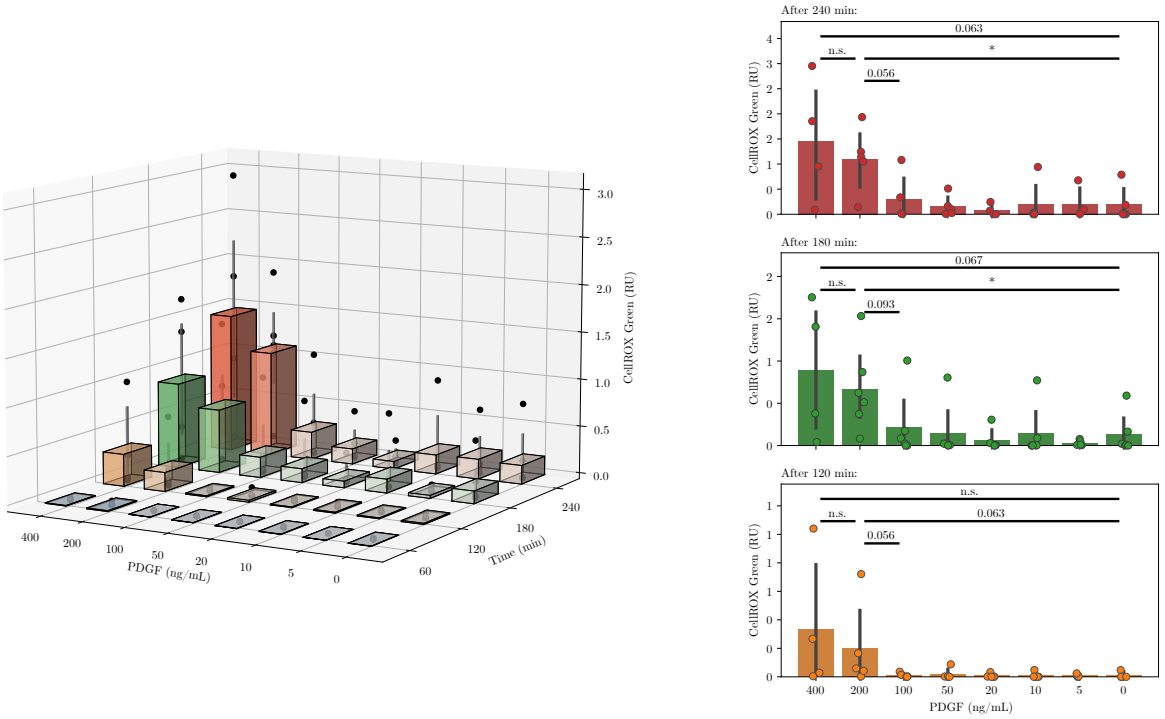

Figure 4.5: PDGF-BB boost titration
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGFβ; followed by two days with IL-1β and PDGF-BB; followed by 4 h boost with 0 - 400 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix. (**A**) 3D visualization: CellROX™ Green signal as a function of PDGF-BB concentration during the boost as well as incubation time. (**B**) 2D visualization: CellROX™ Green signal as a function of PDGF-BB concentration after 120 min, 180 min and 240 min. The shown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by Hoechst 33342 signal. Statistical analysis for (n = 6) biological repeats was performed using Mann-Whitney U test: $* : p < 0.05; ** : p < 0.01$. Note that not every biological repeat covered *all* PDGF-BB concentration.

of PDGF-BB (0 - 400 ng/mL). The signal was detected after 60, 120, 180 and 240 min in HBSS. As seen in figure 4.5, CellROX™ Green signal is negligible after 60 min and then increases with elongated boost times. Moreover, CellROX™ Green signal stays negligible for boost concentrations < 100 ng/mL PDGF-BB. After 180 and 240 min (figure 4.5 B top and middle), CellROX™ Green signal is significantly increased for boost with 200 ng/mL PDGF-BB in comparison to no boost (0 ng/mL PDGF-BB). While the signal in wells boosted with 400 ng/mL PDGF-BB was on average higher than the signal after boost with 200 ng/mL PDGF-BB, this increase was not reproducable. In two repeats the signal was extremely hight, in the other two repeats it collapsed.

Overall, the trend of greatly increased CellROX™ signal for a boost with 100 as well as 200 ng/mL PDGF-BB was consistent within biological repeats; however, the variance between repeats was almost as high as differences between the conditions. Potential causes

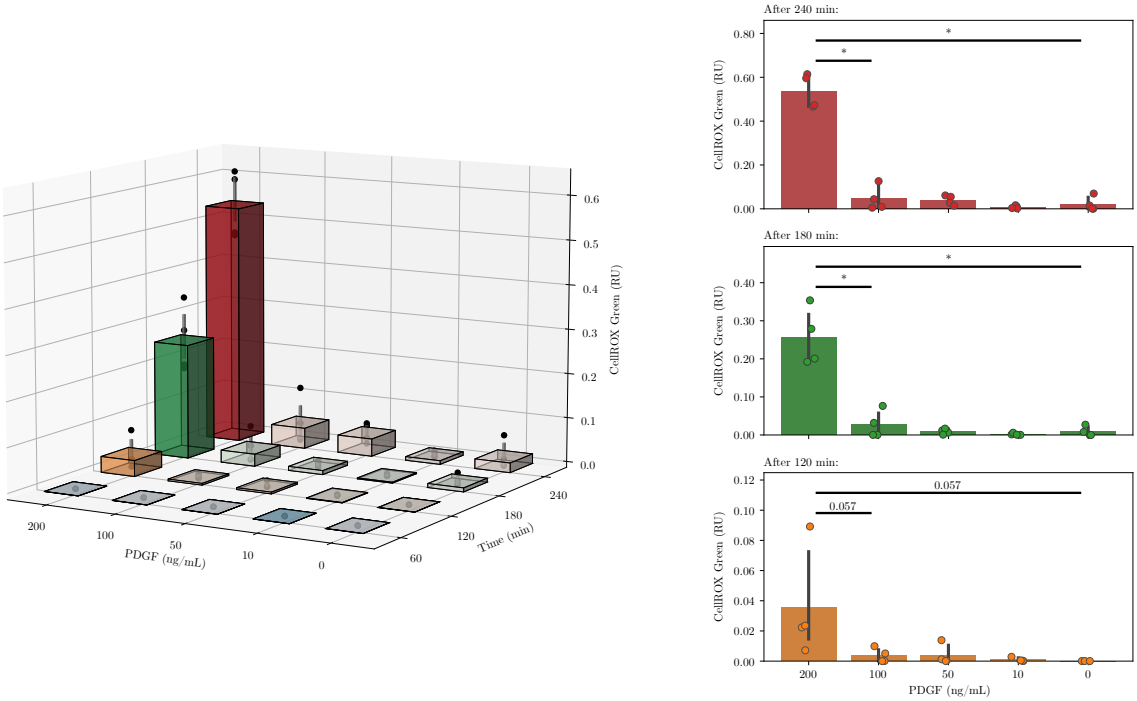

Figure 4.6: PDGF-BB boost titration - normalized
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGFβ; followed by two days with IL-1β and PDGF-BB; follwoed by 4 h boost with 0 - 200 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix. (**A**) 3D visualization: CellROX™ green signal as a function of PDGF-BB concnentration during the boost as well as incubation time. (**B**) 2D visualization: CellROX™ green signal as a function of PDGF-BB concnentration after 120 min, 180 min and 240 min. Shown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by Hoechst 33342 signal, further the signal was normalized via the total signal of the biological repeat. Statistical analysis for (n = 4) biological repeats was performed using Mann-Whitney U test: $* : p < 0.05; ** : p < 0.01$.

for this phenomenon are discussed in section **??**. To account for this large variation between biological repeats, the assay was reevaluated by the selection of shared conditions among the biological repeats normalized to the cumulative intensity of all conditions of the biological repeat (see figure 4.6). This step compensates for differences between biological repeats. The interpretation of the results remain unaffected by normalization: CellROX™ Green signal after 180 and 240 min is significantly higher for cells boosted with 200 ng/mL PDGF-BB than cells that were not boosted (0 ng/mL PDGF-BB).

Rescue of ROS production using NAC

Finally, a rescue experiment was performed to verify that the observed signal in the CellROX™ assay was due to the generation of ROS. For this, ROS generation was quenched by the addition of 2, 4, or 8 mM of NAC. Indeed, a clear trend can be observed: HAoSMCs

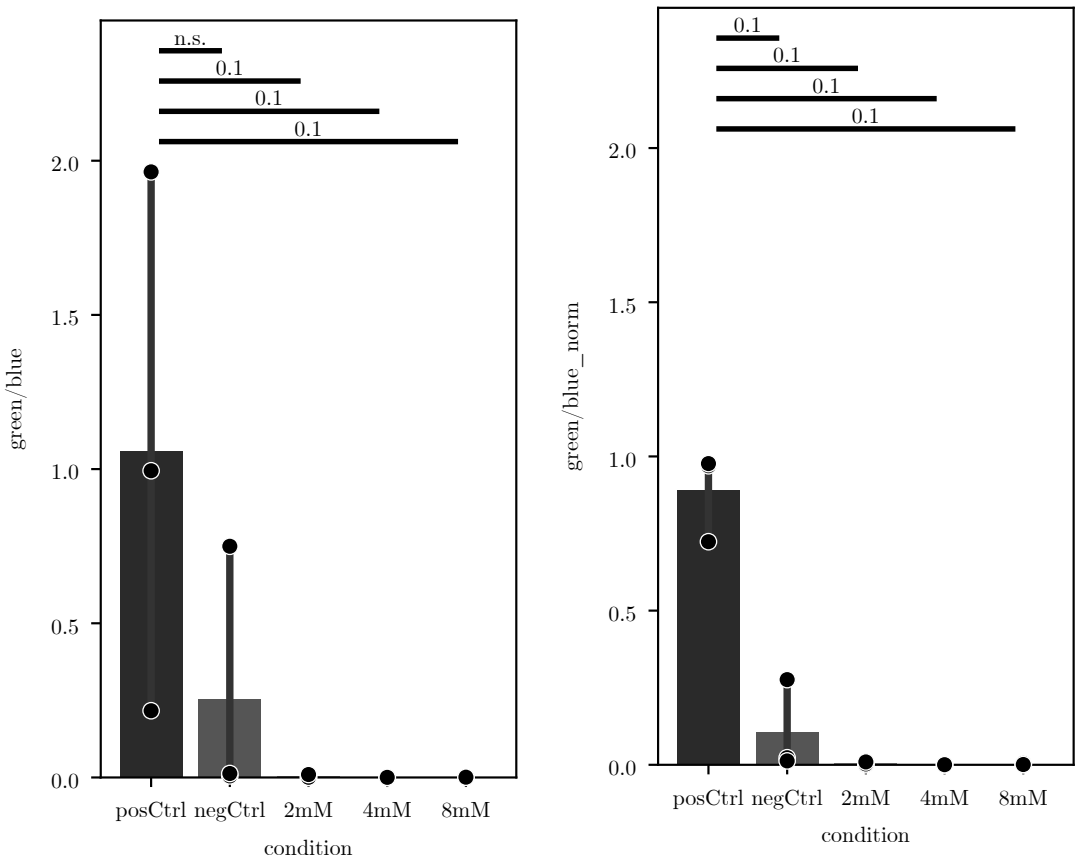

Figure 4.7: ROS generation due to PDGF-BB boost can be rescued with NAC
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGFβ; followed by two days with IL-1β and PDGF-BB; followed by 3 h boost with 200 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix. Cells were treated with 2, 4, or 8 mM of NAC 2 h before the assay. Shown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by Hoechst 33342 signal (**A**), further the signal was normalized via the total signal of the biological repeat (**B**). Statistical analysis for (n = 4) biological repeats was performed using Mann-Whitney U test: $* : p < 0.05; ** : p < 0.01$. pos Ctrl: not treated with NAC, negCtrl: no boost with PDGF-BB

treated with NAC show no signal. However, this trend remained statistically insignificant for triplicates.
It is noteworthy that the signal only builds up over 15 - 20 min under the microscope after the cells were taken out of the incubator. This observation indicates that generation of ROS might not be exclusively triggered by PDGF-BB boost. However, it could also require additional contributors like the loss of the optimized atmosphere of 37°C and 5 % CO_2 in

the incubator. This fact might not have surfaced during the titration assay because cells were taken out of the incubator after one hour to image them for the first time.

4.3 Database and GWAS Visualizer

Curation of Data

Table 4.8: List of Database Tables
List of all the datasets and corresponding tables which were funneled into the database. For primary keys, foreign keys as well as fields on which an idex exists, please consulte figure 4.9. The size of the tables (and accompanying indices) is indicated by the number of databank pages that are reserved for the data, each page fitting 4096 bytes.

Data	Tables	Page count(including indices)
GWAS Summary stats	variation	418,318
	gwas_meta_cad	867,025
	identified_proxy_SNPs_tbl	4
HGNC gene list	hgnc_all_symbols_tbl	826
	hgnc_approved_symbols_tbl	592
Linked SNPs	linked_SNPs_tbl	8,819
	population_tbl	1
	consequence_tbl	1
Ensembl Genome Annotation	ensembl_genelist_tbl	613
	ensembl_genelist_biotypes_tbl	1
Ensembl Regulatory Build	ensembl_reg_build_tbl	8,778
	ensembl_reg_build_features_tbl	1
TSS	tss_tbl	481
Open Target Genetics Scores	opentarget_l2g_tbl	40,984
GWAS catalog	gwascatalog_associations_tbl	10,569
	gwascatalog_studies_tbl	326
TADs	tad_tbl	902
	tad_sample_tbl	1
scATAC seq textcite{}	clint_miller_tbl	12,370
	clint_miller_biotypes_tbl	1
scATAC seq CATlas	catlas_tbl	308,574
	catlas_biotypes_tbl	3
ABC model	abc_tbl	153,920
	abc_targetgenes_tbl	84
	abc_celltypes_tbl	3
	abc_classes_tbl	1
ENCODE cCREs	ENCODE_CCRE	4,451,476
	ENCODE_CCRE_META	107
total	-	6,284,781 (≈ 25.75 GB)

The first step towards visualization of GWA study data and GWA study studies, was the curation of relevant complementary data. Datasets from diverse data sources were downloaded and funneled into an SQLite3 database as described in section 3.5. A structured

query language (SQL) database is a two-dimensional relational database that allows easy and fast access to the data for visualization purposes. The data types and their applications are briefly described in section 1.6. All database tables and their sizes, are summarized in table 4.8. The relationships between the tables and fields serving as a primary key, foreign, or fields on which an index exists, are summarised in the databases entity-relationship (ER) diagram in figure 4.9.

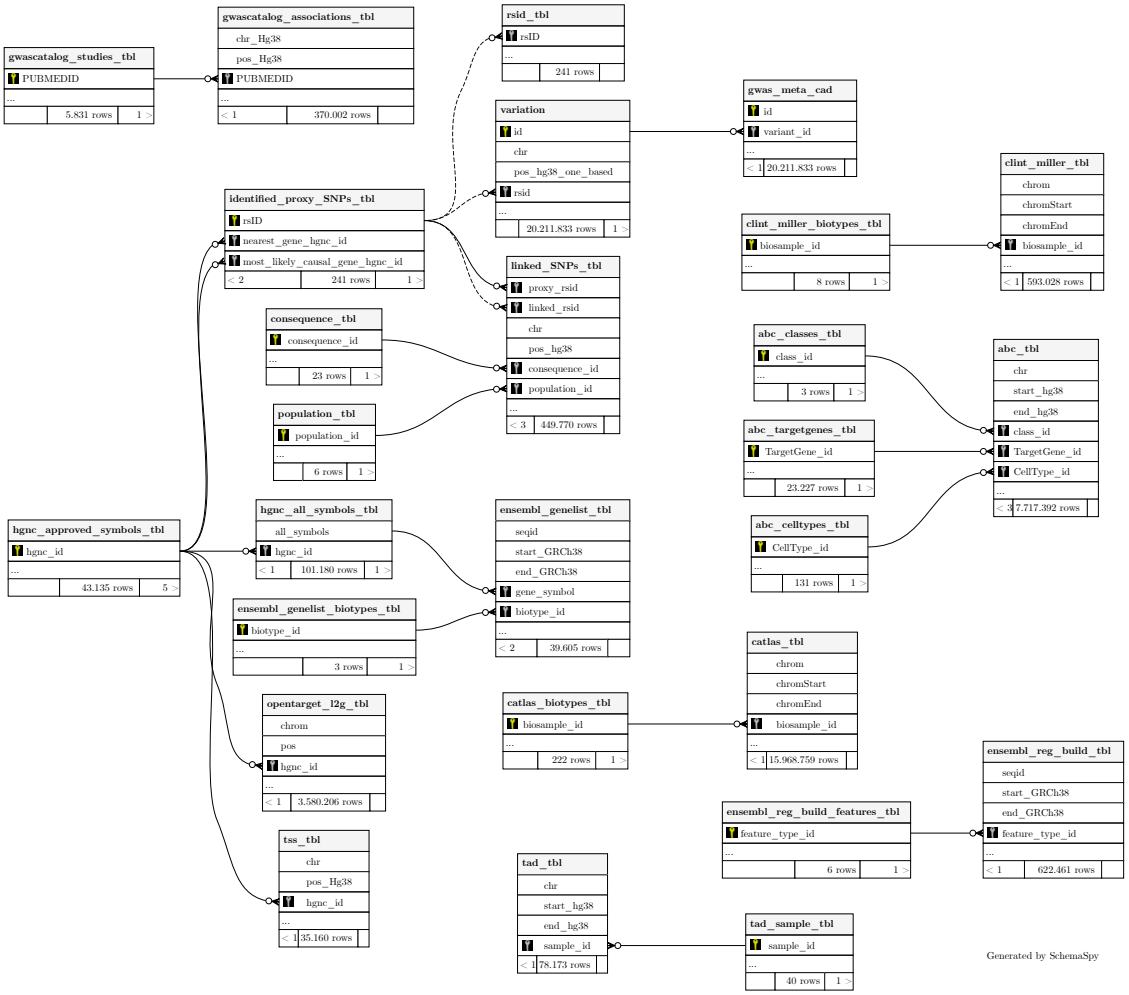

Figure 4.9: Entity-Relationship Diagram of the Database
Fields and relationships of the tables listed in table 4.8. On spelled out columns an index exists or they are primary or forgein keys. The diagram was generated via SchemaSpy.

Figure 4.10: The GWA study Navigator
General content of the GWA study Navigator. The tool contains a manhattan plot with GWA study summary statistics, containing an additional annotation for variants that are in LD with the variant central to the analysis. Further variants identified as proxy variants for other phenotypes are included. Finally, the data are aligned with genomic elements such as genes, regulatory elements, scATAC-seq data as well as the ABC model. More details can be assessed by a hover effect as shown in figure 4.11.

Visualization

Implementing the initially intended use case for the data, a visualization tool for GWA study summary statistics was built according to section 3.6. As shown in figure 4.10, the GWA study Navigator consists of a separate search bar with a field to specifically search for variants by their rsID and a field that allows searching for genes by their symbol. If the searched gene is associated with one of the proxy variants in Aragam et al. (2021), the tool returns a list of these variants. Else the tool returns the most significant variant in the proximity of the searched gene. After a variant was chosen, the tool displays the GWA study summary statistics in a 500 kb window centered around the selected variant to the output panel. GWA study summary statistics are visualized as a zoomed-in Manhatten plot, showing the position of a variant on hg38 on the x-axis and its p-value on the y-axis. r^2 values of variants in LD are color-coded. In addition, the most severe consequence for all linked variants predicted by VEP is indicated by the type of glyph. The minor allele frequency (MAF) and effect size (β) are included in the hover overlay (figure 4.11 A). Below this plot, variant trait associations from the GWA study catalog are indicated for variants that are in LD with the variant central to the analysis (figure 4.11 B). Furthermore, the region is aligned with protein-coding genes, lncRNAs and miRNAs annotated in Ensembl. The names of genes that are associated with the variant of interest (open target genetics L2G score > [FIND THE THRESHHOLD]) are labeled in red. In addition, regulatory elements from the Ensembl regulatory build are displayed. Finally, scATAC-seq data and enhancers-promotor links from the ABC model were aligned, automatically hiding tracks that have no elements in the visualized region.
The GWA study visualizer additionally has a settings tab, in which individual tracks can be hidden.

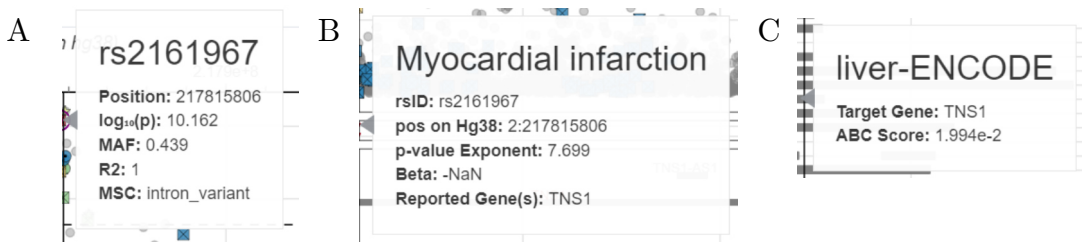

Figure 4.11: The GWA study Visualizer - Hovereffect
Exampalry hover effects for features displayed in the GWA study Navigator. (**A**) Hover for variants in the manhattan plot. (**B**) Hover for variant phenotype associations. (**C**) Hover for cell type specific enhancers in the ABC model.

4.4 Enrichment analysis

The only data that is not displayed in the plot are ENCODE cCREs which were subjected to an enrichment analysis. The annotated biosamples are checked for significant enrichment of cCREs that are overlapping with proxy SNPs identified in the CAD GWA study or variants that are in LD with these ($r^2 > 0.6$). For more details please refer to section 3.7.

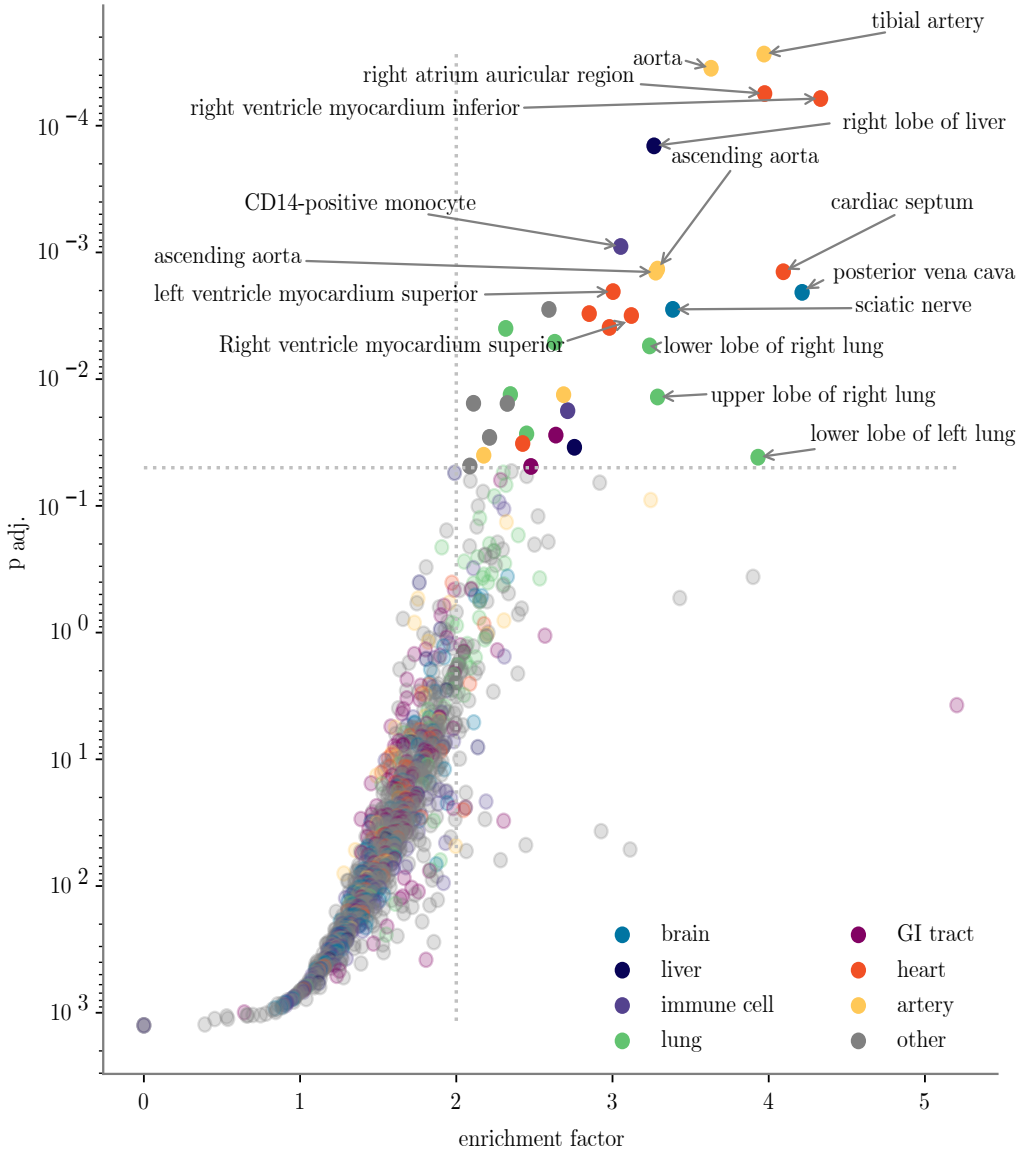

Figure 4.12: Enrichment Analysis for overlap of CAD GWA study proxy variants with tissue specific cCREs
p-values and enrichment factors for the overlap of CAD GWA study proxy variants (and variants in LD) and tissue specific cCREs. For details please refer to section 3.7.

As seen in figure 3.10, statistical significant enrichment ($p_{adj.} < 0.05$) was observed for 34 biosamples. Using the biosample annotations from Cellosaurus, these biosamples were assigned to their tissue of origin. The most prominent groups of origin tissues are the heart (8), the lungs (7), and arteries (6), as summarized in table 4.13. Other tissues included the liver, the gastrointestinal (GI) tract, the brain, and immune cells (CD14+ monocytes).

Table 4.13: Tissues Found in the Enrichment Analysis
Tissues of biosamples which show statitically significant overlap between CAD GWA study proxy variants (and variants in LD) and cCREs.

tissue	count in significant biosamples
heart	8
lung	7
artery	6
liver	2
GI tract	2
brain	2
immune cell	2
other	5
total	34

5

Discussion

5.1 PDGF-BB Signaling Seems to Induce a Synthetic Phenotype in HAoSMCs

The crucial role of vSMCs in atherogenesis has been the subject of intense research for the last few decades (Grootaert and Bennett, 2021; Yap et al., 2021). While it has traditionally been assumed that vSMCs adopt a protective role by stabilizing the arteriogenic plaque. This model is rapidly evolving and starting to consider the existence of a diverse set of dedifferentiated phenotypes (Liu and Gomez, 2019). A central hub of this process is an initial dedifferentiated mesenchymal-like phenotype, that displays a proliferative phenotype and reduced expression of contractile markers (Yap et al., 2021). It is thought to be initiated by the TF KLF4, which induces expression of mesenchymal markers such as stem cell antigen-1 (Sca1) (Yap et al., 2021). Amongst other pathways, expression of KLF4 can be induced by PDGF-BB signaling (Liu et al., 2005) via specificity protein 1 (Sp1) (Deaton et al., 2009). Additionally, PDGF-BB suppresses the contractile phenotype by phosphorylation of ETS like-1 protein (Elk-1) (Wang et al., 2004) as well as the expression of dedicator of cytokinesis 2 (Dock2) (Guo et al., 2015), both of which disrupt myocardin/serum response factor (SRF) mediated expression of contractile genes. The mesenchymal-like phenotype is postulated to serve as a precursor for other dedifferentiated vSMCs phenotypes (Yap et al., 2021).

The contractile expression profile of differentiated vSMCs is constantly maintained by myocardin/SRF signaling (Long et al., 2008) and external stimulation by the extracellular matrix (ECM) and cytokines such as TGFβ (Davis-Dusenbery et al., 2011). HAoSMCs used in this thesis, seem to have initially adopted a dedifferentiated phenotype, characterized by the loss of contractile marker CNN1 (Owens et al., 2004) (figure 4.1 top). When stimulated for two days with TGFβ, HAoSMCs display increased expression of CNN1. Additionally, this phenotype shows a significant decrease in basal mitochondrial respiration, ATP production and maximal respiration (figure 4.3 B top), possibly adapting to the energetic needs of the contractile phenotype, which is considered to be quiescent (Dobnikar et al., 2018). Further simulation for 4 additional days with PDGF-BB & IL-1β, yields a (not quite significant after four biological repeats, $p = 0.087$) drop in CNN1 expression (figure 4.1) as well as a rebound of basal mitochondrial respiration, ATP production as well as maximal respiration to similar levels as for initially dedifferentiated VSMCs (figure 4.3

B bottom).
Another important aspect of phenotypic transition and plaque development is the remodeling of the ECM by MMPs (Johnson, 2017). While not significant in 4 biological repeats, PDGF-BB-induced dedifferentiation seems to increase the expression of MMP9 for HAoSMCs cultivated on col I matrix (figure 4.1 bottom). MMP9 is an important component of atherosclerogenesis (Galis et al., 1994) and a biomarker for advanced atherosclerotic lesions (Langley et al., 2017). The fact that this trend is only observable for cells cultivated on col I (figure 4.1 bottom, $p = 0.063$), underlines the bi-directionality of the ECM-vSMC-interactions and the complexity of vSMC dedifferentiation.
Of course the acpdgf-induced phenotype can not be grasped with only two markers and requires a more indepth analysis.

5.2 CellROX™ Green is Suitable for Assessing ROS Generation in HAoSMCs

Evaluating the response to further stimulation with PDGF-BB, the CellROX™ Assay was able to confirm a result previously observed in the group (unpublished). Stimulation of HAoSMCs cultivated on col I matrix and stimulated for 2 days with TGFβ and 4 days with PDGF-BB & IL-1β, are susceptible to the generation of ROS by PDGF-BB boost (figure 4.4). Further evaluating the limits of the used assay, it is obvious, that a threshold concentration of 200 ng/ml PDGF-BB is required to induce a significant increase in signal over the negative control (0 ng/mL) (figure 4.5). It was further observed, that the signal highly depends on the incubation time. While the trend for each biological repeat is clear, the variance between repeats is almost as high. The assay is working but could greatly benefit from retroactive normalization (figure 4.6) of further optimization towards reproducibility - reducing the required amount of required biological repeats. A potential parameter to explore is the use of different CellROX™ Green concentrations. Finally, a recovery experiment was performed. Before and during the boost, cells were co-incubated with NAC, a potent antioxidant. While not significant after 3 biological repeats, a strong trend was observable, that cells treated with NAC show no CellROX™ Green signal, supporting the expectation that the observed signal is indeed due to the generation of ROS (figure 4.6).
Moreover, it needs to be evaluated if the used PDGF-BB concentration of 200 ng/ml ($\widehat{=}$8.25 nM) is physiologically relevant. Unfortunately, cytokine concentrations are usually assessed as plasma concentrations and no *in vivo* data for local concentrations during paracrine signaling exists. While the manufacturer describes the half maximal effective concentration (EC_{50}) for PDGF-BB-induced proliferation of Balb/c 3T3 cells between 1.0 - 3.0 ng/mL (PeproTech EC Limited, 2022), higher concentrations have frequently been used in the literature. Graves et al. (1996) observed increased formation of cyclic adenosine monophosphate (cAMP) until 10 nM ($\widehat{=}$240 ng/mL) PDGF-BB when assessing the dose-response relationship between cAMP formation after PDGF-BB stimulation of SMCs. Newman et al. (2021b) use 50,ŋ/mL PDGF-BB for the differentiation of murine vSMCs in the context of atherosclerosis, and Bouzigues et al. (2014) identified 100 ng/mL as a saturating concentration for the generation of H_2O_2 as a response to PDGF-BB signaling in vSMCs.
The next up-and-coming experiment would be the rescue experiment to confirm that the

generation of ROS is indeed caused by PDGF-BB stimulation. Namely by the knockdown of the PDGFRβ. The same approach could be pursued to study downstream factors of PDGFR signaling that are involved in the generation of ROS. An exemplary candidate would be STAT1 (STAT1), a TF whichs deletion reduces plaque formation during atherogenesis and is a required component of PDGF-signaling induced inflammation (He et al., 2015). In addition to its genomic function, it can be imported into mitochondria where it interacts with respiratory complexes and triggers the generation of ROS (Wang et al., 2018a) during hepatic (Lee et al., 2007) and interferon (IFN) induced cancer cell apoptosis (Wang et al., 2018a).
Finally, it has to be addressed, that during the recovery experiment with NAC, the CellROX™ Green signal would multiple times only develop only after cells were taken out of the controlled environment of 37°C and 5 %CO_2. This suggests, that the PDGF-BB is not the sole trigger of ROS generation. To follow up on this idea, it would be beneficial to repeat the experiment under better-controlled conditions. We additionally tried to assess oxidative stress with an anti-8-oxoguanine antibody that detects 8-oxoguanine, a base modification caused by ROS. An attempt that unfortunately failed because the cultivation of HAoSMCs for 7 days in M231 + 1 % FBS was sufficient to induce oxidative damage to the genome (results not shown -> maybe I'll still include them).

5.3 The GWAS Navigator

Like all primates, humans are extremely visual creatures. We have evolved specialized brain structures for the processing of visual stimuli (Kaas and Balaram, 2014), granting us superior recognition of visual patterns (Mattson, 2014). Thus, making the tools for visualization of data, powerful and important resources for interactive exploration as well as scientific communication.
The GWA study Navigator was developed to display CAD GWA study summary statistics in an easily accessible format for medical researchers. In an iterative process, a multitude of possible implementation approaches was explored, finally arriving at the prototype presented in this thesis. At this point, the tool is built as a bokeh application (section 3.6) that dynamically fetches data from an SQLite database and renders it to the browser.
Databases are a structured collection of data and a stable of data science, providing many advantages over the storage of data in the form of spreadsheets such as access speed, maintainability, and multiuser access. They are designed to hold large collections of data and provide secure and fast access by querying via specifically designed database engines. Relational databases, like SQLite, are the most popular way of flexible representing data in the form of tables with columns and rows. They are usually queried and manipulated with commands using SQL, an internally consistent, human-readable programming language. (Oracle Corporation, 2022a; Oracle Corporation, 2022b) SQLite is a public domain database engine that generates cross-platform, single file databases and is the most used database engine worldwide (The SQLite Consortium, 2022).
While certainly not the only option, bokeh fulfills all the basic requirements for the task at hand. Combining the elegant visualization resources of rendering data with hypertext markup language (HTML), CSS & JS to the browser with the powerful data processing capabilities of python. All bundled into one easy-to-learn ecosystem, providing a level of abstraction that is required for the construction of a prototype. Additionally, the bokeh

server makes the application easily deployable for potential use on a local network (Bokeh Development Team, 2022).
Overall the GWA study Navigator grants a first glance at the genomic context of disease-associated genomic loci. The next step in its development should undoubtedly be the local deployment for the rest of the lab. It provides basic functionality and the possibility for implementation of many additional features. Reaching from basic improvements to usability in the formed tissue-specific annotations to the displayed tracks and the selection tool, to the expansion with new datasets.

5.4 Overlap of CAD Associated Variants with Regulatory Elements is Enriched in Heart, Artery & Lung Tissue

The database makes all the collected not only easily accessible for visualization purposes, but also follow-up studies. The curated data is utilized in an initial postGWAS analysis, scanning for biosamples with cCREs enriched in CAD GWA study variants via Fisher's exact test. This way identifying 34 biosamples (of [NUMBER] tested) that show significant overrepresentation (figure 4.12 & table 4.13). After annotation of these biosamples, over 40 % (14/34) of enriched biosamples stem from heart or artery tissue and are therefore directly affected by arteriosclerosis. An additional 20 % stem from lung tissue, an observation in line with the often reported association between heart- and lung disease (Carter et al., 2019; Han et al., 2007). The association of heart- and lung disease prevails even after adjustment for shared risk factors such as tobacco or age. Additionally, Au Yeung et al. (2018) were recently able to demonstrate that greater forced expiratory volume in 1 second (FEV_1) decreases the risk of CAD via Mendelian randomization. Still, the causality of this relationship remains unclear. While it is tempting to speculate that impaired lung function or systematic inflammation by chronic diseases like chronic obstructive pulmonary disease (COPD) result in an elevated risk for cardiovascular diseases, such hypotheses are difficult to evaluate due to reverse causation (Nowak, 2018). CAD might also be the risk factor for lung diseases or both pathologies could share additional not properly adjusted confounding factors. Similarly, the identification of lung tissue in our analysis might hint at the involvement of the lung during the development of CAD or a shared genomic predisposition of heart- and lung disease. Following up on the topic of systemic inflammation, the immune cells in which cCREs enrich are CD14+ monocytes (table **??**), a cell type that is known for the secretion of proinflammatory cytokines during injury or inflammation (Kapellos et al., 2019). Interestingly, CD14++CD16+CCR2+ & CD14++CD16-CCR2+ monocytes show significantly higher counts in patients with acute HF over patients with stable HF or CAD (Wrigley et al., 2013).
Finally, the same method and already collected data could be applied to check for the overlap of disease-associated variants with the enhancers identified as part of the ABC model. Further, using the enhancer-promotor connections, to identify potentially affected genes.

Acronyms

24-well multidish	Nunc™Cell-Culture Treated Multidish 24
ABC	activity by contact
API	application programming interface
ATAC-seq	assay for transposase-accessible chromatin using sequencing
ATP	adenosine triphosphate
CAD	coronary artery disease
cAMP	cyclic adenosine monophosphate
cCRE	candidate cis-regulatory element
CNN1	calponin 1
CO_2	carbon dioxide
col I	collagen type I
COPD	chronic obstructive pulmonary disease
Cq	quantification cycle
CRISPR	clustered regularly interspaced short palindromic repeats
CSS	cascading style sheets
CTCF	CCCTC binding factor
cDNA	complementary DNA
DNA	deoxyribonucleic acid
Dock2	dedicator of cytokinesis 2
EC_{50}	half maximal effective concentration
ECAR	extracellular acidification rate
ECM	extracellular matrix
Elk-1	ETS like-1 protein
pELS	proximal enhancer-like elements
dELS	distal enhancer-like elements
ENCODE	ENCyclopedia Of DNA Elements project
eQTL	expression quantitative trait loci
ER	entity-relationship
FBS	fetal bovine serum
FCCP	carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FEV_1	forced expiratory volume in 1 second
FTP	file transfer protocol
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GES analysis	gene set enrichment analysis
GI	gastrointestinal
GWA study	genome wide association study
H3K27ac	histone 3 lysine 27 acetylation

(Continued)

H3K4me3	histone 3 lysine 4 trimethylation
H_20	dihydrogen monoxide (water)
H_2O_2	hydrogen peroxide
HAoSMC	human aortic SMC
HBSS	Hanks balanced salt solution
HF	heart failure
hg19	Genome Reference Consortium Human Build 37
hg38	Genome Reference Consortium Human Build 38
HGNC	Human Gene Nomenclature Consortium
HTML	hypertext markup language
IFN	interferon
IL-1β	interleukin 1 beta
JS	javascript
JSON	JavaScript Object Notation
KLF4	Kruppel-like factor 4
L2G	link to gene
LD	linkage disequilibrium
M231	Human Vascular Smooth Muscle Cell Basal Medium (Medium 231)
MAF	minor allele frequency
MAP	mitogen activated protein
MI	myocardial infarction
MMP	matrix metallopeptidase
MMP9	MMP 9
NAC	N-acetylcysteine
O_2	elemental oxygen
$O_2^{\bullet-}$	superoxide anion radical
OCR	oxygen consumption rate
PBS	phosphate buffered solution
qPCR	quantitative PCR
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PDGF-BB	PDGF-BB
PDGFR	PDGF receptor
PI3K	phosphatidylinositol 3'-kinase
PIP	posterior inclusion probability
PLS	promoter-like elements
REST	representational state transfer
miRNA	micro RNA
lncRNA	long non-coding RNA
mRNA	messenger-RNA
RNA	ribonucleic acid
ROS	reactive oxygen species
RT	reverse transcription
sc	single-cell

(Continued)

Sca1	stem cell antigen-1
SMC	smooth muscle cell
SMGS	Smooth Muscle Cell Growth Supplement
vSMC	vascular SMC
SNP	single nucleotide polymorphism
Sp1	specificity protein 1
SQL	structured query language
SRF	serum response factor
STAT	signal transducers and activators of transcription
STAT1	STAT1
TAD	topologically associated domain
TF	transcription factor
TGFβ	transforming Growth Factor beta
TSS	transcription start sites
USCS	University of California Santa Cruz
VEP	variant effect predictor

Units

d day(s)
h hour(s)
M molar
mg milligram
min minute(s)
mL milliliter
mM milimolar
nM nanomolar
pM picomolar
s second(s)
°C degree celsius
µg microgram
µL microliter
µM micromolar

Bibliography

Disease Control and Prevention, Centers for (2022). *Heart Disease Facts | Cdc.Gov*. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/heartdisease/facts.htm (visited on 06/07/2022).

Fryar, Cheryl D (2012). Prevalence of Uncontrolled Risk Factors for Cardiovascular Disease: United States, 1999–2010, 8.

National Health Service (2017). *Heart Attack*. URL: https://www.nhs.uk/conditions/heart-attack/ (visited on 06/07/2022).

Montalescot, Gilles et al. (2013). 2013 ESC Guidelines on the Management of Stable Coronary Artery Disease: The Task Force on the Management of Stable Coronary Artery Disease of the European Society of Cardiology. European Heart Journal *34*, 2949–3003. DOI: 10.1093/eurheartj/eht296.

Tucker, William D., Arora, Yingyot, and Mahajan, Kunal (2022). 1. 1, Treasure Island (FL): StatPearls Publishing.

Yap, Carmen et al. (2021). Six Shades of Vascular Smooth Muscle Cells Illuminated by KLF4 (Krüppel-Like Factor 4). Arteriosclerosis, Thrombosis, and Vascular Biology, 41:2693–2707. DOI: 10.1161/ATVBAHA.121.316600.

Liu, Mingjun and Gomez, Delphine (2019). Smooth Muscle Cell Phenotypic Diversity. Arteriosclerosis, Thrombosis, and Vascular Biology *39*, 1715–1723. DOI: 10.1161/ATVBAHA.119.312131.

Grootaert, Mandy O J and Bennett, Martin R (2021). Vascular Smooth Muscle Cells in Atherosclerosis: Time for a Re-Assessment. Cardiovascular Research *117*, 2326–2339. DOI: 10.1093/cvr/cvab046.

Goumans, Marie-José and Dijke, Peter ten (2018). TGF- Signaling in Control of Cardiovascular Function. Cold Spring Harbor Perspectives in Biology *10*, a022210. DOI: 10.1101/cshperspect.a022210.

Batlle, Eduard and Massagué, Joan (2019). Transforming Growth Factor- Signaling in Immunity and Cancer. Immunity *50*, 924–940. DOI: 10.1016/j.immuni.2019.03.024.

Davis-Dusenbery, Brandi N. et al. (2011). Down-Regulation of Krüppel-like Factor-4 (KLF4) by MicroRNA-143/145 Is Critical for Modulation of Vascular Smooth Muscle Cell Phenotype by Transforming Growth Factor- and Bone Morphogenetic Protein 4. The Journal of Biological Chemistry *286*, 28097–28110. DOI: 10.1074/jbc.M111.236950.

Takahashi, Kazutoshi et al. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872. DOI: 10.1016/j.cell.2007.11.019.

Pan, Huize et al. (2020). Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. Circulation *142*, 2060–2075. DOI: 10.1161/CIRCULATIONAHA.120.048378.

Chen, Po-Han, Chen, Xiaoyan, and He, Xiaolin (2013). Platelet-Derived Growth Factors and Their Receptors: Structural and Functional Perspectives. Biochimica et biophysica acta *1834*, 2176–2186. DOI: 10.1016/j.bbapap.2012.10.015.

Heldin, Carl-Henrik (2013). Targeting the PDGF Signaling Pathway in Tumor Treatment. Cell Communication and Signaling *11*, 97. DOI: 10.1186/1478-811X-11-97.

Hu, Weining and Huang, Yu (2015). Targeting the Platelet-Derived Growth Factor Signalling in Cardiovascular Disease. Clinical and Experimental Pharmacology and Physiology *42*, 1221–1224. DOI: 10.1111/1440-1681.12478.

Andrae, Johanna, Gallini, Radiosa, and Betsholtz, Christer (2008). Role of Platelet-Derived Growth Factors in Physiology and Medicine. Genes & Development *22*, 1276–1312. DOI: 10.1101/gad.1653708.

Levéen, P. et al. (1994). Mice Deficient for PDGF B Show Renal, Cardiovascular, and Hematological Abnormalities. Genes & Development *8*, 1875–1887. DOI: 10.1101/gad.8.16.1875.

Robson, M. C. et al. (1992). Platelet-Derived Growth Factor BB for the Treatment of Chronic Pressure Ulcers. Lancet (London, England) *339*, 23–25. DOI: 10.1016/0140-6736(92)90143-q.

Raines, Elaine W (2004). PDGF and Cardiovascular Disease. Cytokine & growth factor reviews *15*, 237–254. DOI: 10.1016/j.cytogfr.2004.03.004.

He, Chaoyong et al. (2015). PDGFR Signalling Regulates Local Inflammation and Synergizes with Hypercholesterolaemia to Promote Atherosclerosis. Nature Communications *6* (1), 7770. DOI: 10.1038/ncomms8770.

Newman, Alexandra A. C. et al. (2021a). Multiple Cell Types Contribute to the Atherosclerotic Lesion Fibrous Cap by PDGFR and Bioenergetic Mechanisms. Nature Metabolism *3* (2), 166–181. DOI: 10.1038/s42255-020-00338-8.

Sies, Helmut and Jones, Dean P. (2020). Reactive Oxygen Species (ROS) as Pleiotropic Physiological Signalling Agents. Nature Reviews Molecular Cell Biology *21* (7), 363–383. DOI: 10.1038/s41580-020-0230-3.

Nayernia, Zeynab, Jaquet, Vincent, and Krause, Karl-Heinz (2014). New Insights on NOX Enzymes in the Central Nervous System. Antioxidants & Redox Signaling *20*, 2815–2837. DOI: 10.1089/ars.2013.5703.

Sundaresan, M. et al. (1995). Requirement for Generation of H2O2 for Platelet-Derived Growth Factor Signal Transduction. Science (New York, N.Y.) *270*, 296–299. DOI: 10.1126/science.270.5234.296.

Bouzigues, Cedric I. et al. (2014). Regulation of the ROS Response Dynamics and Organization to PDGF Motile Stimuli Revealed by Single Nanoparticle Imaging. Chemistry & Biology *21*, 647–656. DOI: 10.1016/j.chembiol.2014.02.020.

Uffelmann, Emil et al. (2021). Genome-Wide Association Studies. Nature Reviews Methods Primers *1* (1), 1–21. DOI: 10.1038/s43586-021-00056-9.

Flint, Jonathan (2013). GWAS. Current Biology *23*, R265–R266. DOI: 10.1016/j.cub.2013.01.040.

Schaid, Daniel J., Chen, Wenan, and Larson, Nicholas B. (2018). From Genome-Wide Associations to Candidate Causal Variants by Statistical Fine-Mapping. Nature reviews. Genetics *19*, 491–504. DOI: 10.1038/s41576-018-0016-z.

Lichou, Florence and Trynka, Gosia (2020). Functional Studies of GWAS Variants Are Gaining Momentum. Nature Communications *11*, 6283. DOI: 10.1038/s41467-020-20188-y.

Slatkin, Montgomery (2008). Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. Nature reviews. Genetics *9*, 477–485. DOI: 10.1038/nrg2361.

Mountjoy, Edward et al. (2021). An Open Approach to Systematically Prioritize Causal Variants and Genes at All Published Human GWAS Trait-Associated Loci. Nature Genetics *53* (11), 1527–1533. DOI: 10.1038/s41588-021-00945-5.

Zerbino, Daniel R. et al. (2015). The Ensembl Regulatory Build. Genome Biology *16*, 56. DOI: 10.1186/s13059-015-0621-5.

Moore, Jill E. et al. (2020). Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. Nature *583* (7818), 699–710. DOI: 10.1038/s41586-020-2493-4.

Buenrostro, Jason D. et al. (2013). Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-binding Proteins and Nucleosome Position. Nature Methods *10* (12), 1213–1218. DOI: 10.1038/nmeth.2688.

Buenrostro, Jason D. et al. (2015a). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current Protocols in Molecular Biology *109*, 21.29.1–21.29.9. DOI: 10.1002/0471142727.mb2129s109.

Buenrostro, Jason D. et al. (2015b). Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation. Nature *523* (7561), 486–490. DOI: 10.1038/nature14590.

Fulco, Charles P. et al. (2019). Activity-by-Contact Model of Enhancer–Promoter Regulation from Thousands of CRISPR Perturbations. Nature Genetics *51* (12), 1664–1669. DOI: 10.1038/s41588-019-0538-0.

Nasser, Joseph et al. (2021). Genome-Wide Enhancer Maps Link Risk Variants to Disease Genes. Nature *593* (7858), 238–243. DOI: 10.1038/s41586-021-03446-x.

Lieberman-Aiden, Erez et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289–293. DOI: 10.1126/science.1181369.

Wit, Elzo de and Laat, Wouter de (2012). A Decade of 3C Technologies: Insights into Nuclear Organization. Genes & Development *26*, 11–24. DOI: 10.1101/gad.179804.111.

Dixon, Jesse R. et al. (2012). Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. Nature *485* (7398), 376–380. DOI: 10.1038/nature11082.

Wang, Yanli et al. (2018b). The 3D Genome Browser: A Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions. Genome Biology *19*, 151. DOI: 10.1186/s13059-018-1519-9.

Pombo, Ana and Dillon, Niall (2015). Three-Dimensional Genome Architecture: Players and Mechanisms. Nature Reviews Molecular Cell Biology *16* (4), 245–257. DOI: 10.1038/nrm3965.

Burtenshaw, Denise et al. (2019). Reactive Oxygen Species (ROS), Intimal Thickening, and Subclinical Atherosclerotic Disease. Frontiers in Cardiovascular Medicine *6*, 89. DOI: 10.3389/fcvm.2019.00089.

Aragam, Krishna G. et al. (2021). Discovery and Systematic Characterization of Risk Variants and Genes for Coronary Artery Disease in over a Million Participants, 2021.05.24.21257377. DOI: 10.1101/2021.05.24.21257377.

Huggett, Jim and Bustin, Stephen A. (2011). Standardisation and Reporting for Nucleic Acid Quantification. Accreditation and Quality Assurance *16*, 399. DOI: 10.1007/s00769-011-0769-y.

AgilentTechnologies (2022). *How Agilent Seahorse XF Analyzers Work | Agilent*. URL: https://www.agilent.com/en/products/cell-analysis/how-seahorse-xf-analyzers-work (visited on 06/05/2022).

Thermo Fisher Scientific Inc. (2022). *CellROX Green Reagent, for Oxidative Stress Detection*. URL: https://www.thermofisher.com/order/catalog/product/C10444 (visited on 06/05/2022).

Newman, Alexandra A. C. et al. (2021b). Multiple Cell Types Contribute to the Atherosclerotic Lesion Fibrous Cap by PDGFR and Bioenergetic Mechanisms. Nature Metabolism *3* (2), 166–181. DOI: 10.1038/s42255-020-00338-8.

Bokeh Development Team (2022). *Bokeh: Python Library for Interactive Visualization*.

Langtangen, Hans Peter and Johansen, Anders E. (2015). *Using Web Frameworks for Scientific Applications*. URL: http://hplgit.github.io/web4sciapps/doc/pub/web4sa_plain_all.html#wf:bokeh:flask (visited on 07/14/2022).

Tipney, Hannah and Hunter, Lawrence (2010). An Introduction to Effective Use of Enrichment Analysis Software. Human Genomics *4*, 202. DOI: 10.1186/1479-7364-4-3-202.

Liu, Yan et al. (2005). Kruppel-like Factor 4 Abrogates Myocardin-induced Activation of Smooth Muscle Gene Expression *. Journal of Biological Chemistry *280*, 9719–9727. DOI: 10.1074/jbc.M412862200.

Deaton, Rebecca A., Gan, Qiong, and Owens, Gary K. (2009). Sp1-Dependent Activation of KLF4 Is Required for PDGF-BB-induced Phenotypic Modulation of Smooth Muscle. American Journal of Physiology. Heart and Circulatory Physiology *296*, H1027–1037. DOI: 10.1152/ajpheart.01230.2008.

Wang, Zhigao et al. (2004). Myocardin and Ternary Complex Factors Compete for SRF to Control Smooth Muscle Gene Expression. Nature *428* (6979), 185–189. DOI: 10.1038/nature02382.

Guo, Xia et al. (2015). Dedicator of Cytokinesis 2, A Novel Regulator for Smooth Muscle Phenotypic Modulation and Vascular Remodeling. Circulation Research *116*, e71–e80. DOI: 10.1161/CIRCRESAHA.116.305863.

Long, Xiaochun et al. (2008). Myocardin Is Sufficient for a Smooth Muscle-Like Contractile Phenotype. Arteriosclerosis, Thrombosis, and Vascular Biology *28*, 1505–1510. DOI: 10.1161/ATVBAHA.108.166066.

Owens, Gary K., Kumar, Meena S., and Wamhoff, Brian R. (2004). Molecular Regulation of Vascular Smooth Muscle Cell Differentiation in Development and Disease. Physiological Reviews *84*, 767–801. DOI: 10.1152/physrev.00041.2003.

Dobnikar, Lina et al. (2018). Disease-Relevant Transcriptional Signatures Identified in Individual Smooth Muscle Cells from Healthy Mouse Vessels. Nature Communications *9* (1), 4567. DOI: 10.1038/s41467-018-06891-x.

Johnson, Jason L. (2017). Metalloproteinases in Atherosclerosis. European Journal of Pharmacology *816*, 93–106. DOI: 10.1016/j.ejphar.2017.09.007.

Galis, Z. S. et al. (1994). Increased Expression of Matrix Metalloproteinases and Matrix Degrading Activity in Vulnerable Regions of Human Atherosclerotic Plaques. The Journal of Clinical Investigation *94*, 2493–2503. DOI: 10.1172/JCI117619.

Langley, Sarah R. et al. (2017). Extracellular Matrix Proteomics Identifies Molecular Signature of Symptomatic Carotid Plaques. The Journal of Clinical Investigation *127*, 1546–1560. DOI: 10.1172/JCI86924.

PeproTech EC Limited (2022). *Recombinant Human PDGF-BB*. PeproTech. URL: https://www.peprotech.com/recombinant-human-pdgf-bb (visited on 07/14/2022).

Graves, L. M. et al. (1996). Platelet-Derived Growth Factor Stimulates Protein Kinase A through a Mitogen-Activated Protein Kinase-Dependent Pathway in Human Arterial Smooth Muscle Cells. The Journal of Biological Chemistry *271*, 505–511. DOI: 10.1074/jbc.271.1.505.

Wang, Yan et al. (2018a). The STAT-ROS Cycle Extends IFN-induced Cancer Cell Apoptosis. International Journal of Oncology *52*, 305–313. DOI: 10.3892/ijo.2017.4196.

Lee, Hyun Jung et al. (2007). The Role of STAT1/IRF-1 on Synergistic ROS Production and Loss of Mitochondrial Transmembrane Potential during Hepatic Cell Death Induced by LPS/d-GalN. Journal of Molecular Biology *369*, 967–984. DOI: 10.1016/j.jmb.2007.03.072.

Kaas, Jon H and Balaram, Pooja (2014). Current Research on the Organization and Function of the Visual System in Primates. Eye and Brain *6* (Suppl 1), 1–4. DOI: 10.2147/EB.S64016.

Mattson, Mark P. (2014). Superior Pattern Processing Is the Essence of the Evolved Human Brain. Frontiers in Neuroscience *8*, 265. DOI: 10.3389/fnins.2014.00265.

Oracle Corporation (2022a). *What Is a Database?* URL: https://www.oracle.com/database/what-is-database/ (visited on 07/13/2022).

Oracle Corporation (2022b). *What Is a Relational Database | Oracle*. URL: https://www.oracle.com/database/what-is-a-relational-database/ (visited on 07/13/2022).

The SQLite Consortium (2022). *About SQLite*. URL: https://www.sqlite.org/about.html (visited on 07/13/2022).

Carter, Paul et al. (2019). Association of Cardiovascular Disease With Respiratory Disease. Journal of the American College of Cardiology *73*, 2166–2177. DOI: 10.1016/j.jacc.2018.11.063.

Han, MeiLan K. et al. (2007). Pulmonary Diseases and the Heart. Circulation *116*, 2992–3005. DOI: 10.1161/CIRCULATIONAHA.106.685206.

Au Yeung, Shiu Lun, Borges, Maria-Carolina, and Lawlor, Debbie A. (2018). Association of Genetic Instrumental Variables for Lung Function on Coronary Artery Disease Risk. Circulation: Genomic and Precision Medicine *11*, e001952. DOI: 10.1161/CIRCGEN.117.001952.

Nowak, Christoph (2018). Lung Function and Coronary Artery Disease Risk. Circulation: Genomic and Precision Medicine *11*, e002137. DOI: 10.1161/CIRCGEN.118.002137.

Kapellos, Theodore S. et al. (2019). Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases. Frontiers in Immunology *10*.

Wrigley, Benjamin J. et al. (2013). CD14++CD16+ Monocytes in Patients with Acute Ischaemic Heart Failure. European Journal of Clinical Investigation *43*, 121–130. DOI: 10.1111/eci.12023.