



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

Identifikation genetischer Risikovarianten für Atherosklerose via oxidativem Stress Assay in glatten Gefäßmuskelzellen 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 26. 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

Nach Angaben der Weltgesundheitsorganisation (WHO), sind „Herz-Kreislauf-Erkrankungen (KHK) [...] die weltweit häufigste Todesursache und fordern jedes Jahr schätzungsweise 17,9 Millionen Todesopfer.“ (WHO, 2022). Ein entscheidender Schritt für die Entwicklung erfolgreicher Therapieansätze ist die Aufklärung der zugrunde liegenden molekularen Prozesse.

Glatten Gefäßmuskelzellen (gGMZ) spielen eine wichtige (Doran u. a., 2008), aber nur unvollständig verstandene Rolle in der Entstehung und dem Erhalt von atherosklerotischen Plaques (Liu und Gomez, 2019; Grootaert und Bennett, 2021; Yap u. a., 2021). Wir stellen die Hypothese auf, dass die anhaltende Stimulation von gGMZ mit proinflammatorischen Zytokinen, wie z. B. PDGF-BB, oxidativen Stress verursacht und so zur Pathogenese beiträgt.

Wir kommen zu dem Schluss, dass das von uns genutzte CellROX™ Assay eine geeignete Methode zur Evaluation dieser Hypothese ist und bereiten Vorschläge für weitere Optimierungen des Assays. Zusätzlich zeigen wir, dass die Inkubation von *in vitro* dedifferenzierten HAoSMCs mit hohen Konzentrationen an PDGF-BB die Produktion von reaktiven Sauerstoffspezies (ROS) induziert.

Ergänzt zu diesem Hypothesen-getriebenen Ansatz konsultieren wir Daten aus Hochdurchsatz-Verfahren in Form von Genomweite Assoziationsstudien (GWAS). GWAS sind ein Kernstück der modernen Genetik und eröffnen grandiose Möglichkeiten zur Identifikation krankheitsassoziierter Loci (Uffelmann u. a., 2021). Für ein besseres Verständnis des genomischen Kontexts von KHK Risikovarianten präsentieren wir den GWAS Navigator - eine web basierte, interaktive Anwendung, die GWAS Daten des CARDIoGRAMplusC4D Konsortiums (Aragam u. a., 2021) in Verbindung mit kuriierten regulatorischen Elementen visualisiert.

Abschließend nutzen wir die gesammelten Daten für eine Anreicherungsanalyse, in der wir für die Überschneidung von KHK Risikovarianten mit gewebe- oder zellspezifischen regulatorischen Elementen testen. Wir beobachten eine signifikante Überrepräsentation solcher Überschneidungen vor allem in Proben des Herzens, der Lunge und von arteriellem Gewebe. Ein Ergebnis, das die oft beschriebene Assoziation von Herz- mit Lungenerkrankungen unterstreicht (Carter u. a., 2019; Nowak, 2018; Han u. a., 2007).

Abstract

“Cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year.”, according to the world health organization (WHO) (WHO, 2022). The most common CVD is coronary artery disease (CAD) (Centers for Disease Control and Prevention, 2022), which poses a significant challenge for current and future medical professionals. Investigating the molecular basics of this pathological process is pivotal for developing suitable treatment options.

While vascular smooth muscle cell (SMC)s (vSMCs) are essential in initiating and maintaining atherosclerotic plaques (Doran et al., 2008), the exact model for their precise role and function is frequently updated (Liu and Gomez, 2019; Grootaert and Bennett, 2021; Yap et al., 2021). We hypothesize that extensive stimulation of vSMCs with pro-inflammatory cytokines such as platelet-derived growth factor (PDGF)-BB (PDGF-BB) can cause oxidative stress and promote disease progression.

For this, we recapitulated that PDGF-BB boost of *in vitro* dedifferentiated human aortic SMC (HAoSMC) induces the generation of reactive oxygen species (ROS). Further, we tested the limits of the CellROX™ Assay to evaluate this interaction. We conclude that the assay provides a solid basis for analyzing ROS generation due to PDGF-BB stimulation. Additionally, we suggest optimizations to the assay to improve its reproducibility.

In contrast to this hypothesis-driven approach, genome wide association studies (GWAS) are a tremendous observational tool for identifying exciting research targets (Uffelmann et al., 2021). To complement CARDIoGRAMplusC4D GWAS summary statistics (Aragam et al., 2021), we curated and funneled genomic annotations of regulatory elements into a database. This database serves as the foundation for the GWAS Navigator, an interactive web-based visualization tool. It provides easy access to the GWAS summary statistics and their genomic context.

Finally, we integrated the curated data into an enrichment analysis — testing for overlap of CARDIoGRAMplusC4D risk variants with tissue- and cell-specific regulatory elements. We found an overrepresentation of overlaps mainly in biosamples from the heart, arteries, or lungs — a result that underlines the frequently described connection between heart- and lung diseases (Carter et al., 2019; Nowak, 2018; Han et al., 2007).

Acknowledgements

First and foremost, I'd like to thank Dr. Tobias Reinberger for his outstanding supervision. Thank you for always making time for my questions, great advice, motivating words, and genuine interest in my project. Further, I'd like to thank Prof. Dr. Jeanette Erdmann for giving me the chance to be a part of her lab and for her excellent feedback — I truly learned a lot! I'd like to thank Lisa Diner for taking care of my cells when I wasn't available and answering my frequent questions. Of course, I'd like to thank all the other group members. Thank you for the welcoming atmosphere, delicious cake, necessary breaks, and excellent scientific advice!

In addition, this thesis marks the end of my studies at the University of Lübeck. I'd like to thank all the friends who supported me during this time. Those who are becoming incredible scientists themselves and those who enrich my life by having totally different interests. I value your opinions very much, and thank you for taking my mind off my studies sometimes. I appreciate every single one of you; you made Lübeck feel like home.

Finally, I'd like to thank my parents for their immense support. I love you.

Contents

1	Introduction	1
1.1	Coronary artery disease	1
1.2	VSMCs in CAD	1
1.3	TGF β signaling	2
1.4	PDGF signaling	2
1.5	GWAS	3
1.6	Complementary high through put methods	5
1.7	Aim of the thesis	7
2	Material	9
2.1	Manufactors	9
2.2	Cell lines	10
2.3	Kits	11
2.4	Oligonucleotides	11
2.5	Chemicals	11
2.6	Media and supplements	12
2.7	Solutions	12
2.8	Consumables	13
2.9	Devices	14
2.10	Programs and Modules	15
3	Methods	17
3.1	Cultivation and differentiation of HAoSMCs	17
3.2	mRNA quantification	18
3.3	Energy profiling	20
3.4	CellROX TM assay	21
3.5	Curation of data for postGWAS analyses	23
3.6	Visualization of GWAS data	24
3.7	Enrichment analysis	25
4	Results	28
4.1	Differentitaion	28
4.2	Evaluation of oxidative Stress	32
4.3	Database and GWAS Navigator	36
4.4	Enrichment analysis	39

5	Discussion	43
5.1	PDGF-BB signaling induces a synthetic phenotype in HAoSMCs	43
5.2	CellROX™ Green is suitable for assessing ROS generation in HAoSMCs . .	44
5.3	The GWAS Navigator	45
5.4	Overlap of CAD associated variants with regulatory elements is enriched in heart, artery, and lung tissue	46
	Bibliography	48
	Abbreviations and units	55
	Supplement	58
S.1	CellROX™ assay	58
S.2	Evaluation of oxidative stress with an anti-8-oxoguanine antibody	58

1

Introduction

1.1 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 (Centers for Disease Control and Prevention, [2022](#); Fryar, [2012](#)). CAD is characterized by the build-up of fatty plaques in 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 heart's 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 vSMCs. vSMCs 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](#))

Cell types commonly associated with atherogenesis are endothelial cells, immune cells, and vSMCs (Tabas et al., [2015](#)). 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 atherosclerosis is rapidly evolving, and the underlying models are being adjusted accordingly. The black and white idea of vSMCs in atherosclerosis 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 (TGF β) and platelet-derived growth factor (PDGF)-BB (PDGF-BB).

1.3 TGF β signaling

TGF β signaling in general

The term Transforming Growth Factor beta (TGF β) describes a superfamily of cytokines, the most prominent of which is TGF β 1. After secretion and activation, the active TGF β dimer binds to a heteromeric receptor complex. The intracellular signaling is mainly implemented via Smad transcription factors. The cellular effects of TGF β are highly dependent on the cell type and can even be pleiotropic for cells of the same type. The most prominent function of TGF β is its role in the anti-inflammatory regulation of immune cells. (Goumans and Dijke, 2018; Batlle and Massagué, 2019)

TGF β signaling in VSMCs and atherosclerosis

In the context of vSMCs, TGF β promotes proliferation and hypertrophy. Further, it promotes vSMC differentiation via elevated gene expression of contractile marker genes. Additionally, TGF β 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, TGF β 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, and PDGF-DD) have been identified as dimeric combinations 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 PDGF isoform predominantly 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 PDGFR β 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 PDGFR β signaling in pathologic processes like cancer or cardiovascular disease has been a subject of thorough study for decades (Heldin, 2013; Raines, 2004).

PDGF signaling in vSMCs and atherosclerosis

Similarly to the overall role of vSMCs in atherosclerosis, the role of PDGF-BB is still the subject of intensive 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 atherosclerotic 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 PDGFR β 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 PDGFR β is required for vSMC involvement in atherosclerotic 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 atherosclerotic lesions.

ROS in PDGF signaling

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) ROS as intracellular messengers predominantly implement their effect by oxidation protein targets (Zeida et al., 2019).

Human cells contain dozens of proteins 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 secondary 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

GWAS

A fantastic resource for obtaining a first glance at these interactions are GWAS, a method that enables the identification of genetic variants associated with a phenotype. Variants most commonly analyzed in GWAS are single nucleotide polymorphisms (SNPs).

While GWAS 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 GWAS is the first important step in determining causal variants for a phenotype and, therefore, a glimpse into its molecular basis (Uffelmann et al., 2021).

postGWAS

Unfortunately, GWAS 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) are essential to convert the first list of 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 GWAS 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 probabilities (PIPs) that form a credible set of potentially causal variants. Noteworthy, fine-mapping methods are continuously refined and evolve alongside the increasing complexity of the studied phenotypes. Finally, fine-mapping is a purely statistical approach that can only identify potential causal variants which need to be confirmed via complementary approaches. (Schaid et al., 2018; Uffelmann et al., 2021)

After fine-mapping, the following steps aim to gain information on the causal variants' 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) data (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 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).

GWAS and CAD

The history of GWAS for CAD research started in 2007 with the identification of the strongest CAD risk locus on chromosome 9p21 by multiple independent studies (Holdt and Teupser, 2012; Erdmann et al., 2018). Since then, the use of GWAS in research has continuously evolved alongside meaningful advances in sequencing technologies and data availability (Erdmann et al., 2018). This way, incrementally identifying novel variants and arriving at the observation that a large part of CAD risk can be traced to common variants with small effect sizes (Khera and Kathiresan, 2017; Erdmann et al., 2018).

The latest 15 years of research converge in a meta-analysis by the CARDIoGRAMplusC4D Consortium. The analysis combines the results of ten published studies, achieving a sample size of >180,000 cases and almost 1,000,000 control patients. It ultimately yielded a list of 241 associations. Additional postGWAS analyses provide a set of 185 candidate causal genes, serving as a great catalog of potential research targets (Aragam et al., 2021). Nevertheless, GWAS for CAD suffers from similar problems as GWAS in general, and the underlying mechanisms of the disease remain largely unknown. Aragam et al. (2021) nicely summarize: "the translation of statistical associations into actionable biology and potential therapeutic targets requires elucidation of causal variants, genes and intermediate pathways, which has lagged behind the rapid growth in genetic association discoveries."

1.6 Complementary high through put methods

The development of high through put methods combined with the significant increase in computational power over the last few years have paved the way for postGWAS. 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 because identified 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 GWAS loci to their target genes. The computed L2G scores are calibrated to represent a probability (0,1). (Mountjoy et al., 2021)

Ensembl 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 sites. The build considers promoters, 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 promoters
- 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 transposase 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 and 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. At the same time, the edges enrich housekeeping genes and CTCF binding sites, 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 staple of modern bioscience and a great resource. The study by Aragam et al. (2021) provides a vast and important dataset for CAD research.

1 Introduction

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 GWAS Navigator will visualize GWAS summary statistics with different annotations in the form of associated phenotypes from other GWAS and putative regulatory elements.

- Finally, the data curated for the GWAS 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 Manufacturers

Manufacturer	Seat
Agilent Technologies, Inc.	Santa Clara, CA, USA
Applied Biosystems (Brand of LUMITOS AG)	Warrington, UK
Glaswarenfabrik Karl Hecht GmbH & Co. KG	Sondheim vor der Rhön, DE
J.T.Baker® (Brand of Thermo Fisher Scientific Inc.)	Schwerte, DE
Becton Dickinson GmbH	Heidelberg, DE
Bio&SELL GmbH	Feucht, DE
Brand GmbH & Co. KG	Wertheim, DE
Cell Applications, Inc	San Diego, CA, USA
ChemoMetec A/S	Allerød, DK
Eppendorf SE	Hamburg, DE
Eurofins Genomics Germany GmbH	Ebersberg, DE
GFL mbH	Burgwedel, DE
Gibco® BRL (Brand of Thermo Fisher Scientific Inc.)	Gaithersburg, MD, USA
GitHub, Inc. (Part of Microsoft Corporation)	San Francisco, CA, USA
Hanna Instruments Deutschland GmbH	Vöhringen, DE
Heidolph Instruments Labortechnik	Schwabach, DE
Heraeus Holding GmbH	Hanau, DE

Continued on next page

(Continued)

Manufacturer	Seat
ibidi GmbH	Gräfelfing, DE
Invitrogen™ AG (Brand of Thermo Fisher Scientific Inc.)	Schwerte, DE
JetBrains s.r.o.	Prague, CZ
Keyence Corporation	Osaka, JP
Kisker Biotech GmbH & Co. KG	Steinfurt, DE
Lonza Group AG	Basel, CHE
Mettler-Toledo GmbH	Gießen, DE
Merck KGaA	Darmstadt, DE
Microsoft Deutschland GmbH	München, DE
New England Biolabs GmbH	Ipswich, MA, USA
Nikon Corporation	Minato, JP
Pechiney Plastic Packaging, Inc.	Chicago, IL, USA
PeproTech® (Brand of Thermo Fisher Scientific Inc.)	Hamburg, DE
Python Software Foundation	Beaverton, OR, USA
Sarstedt AG & Co.	Nürnberg, DE
Sartorius AG	Göttingen, DE
SensoQuest GmbH	Göttingen, DE
Serif (Europe) Ltd.	Nottingham, UK
Sigma-Aldrich Co. LLC.	St. Louis, MO, USA
The SQLite Consortium (Contact via Hipp, Wyrick & Company)	Charlotte, NC, USA
Thermo Fisher Scientific Inc.	Waltham, MA, USA
ZIEGRA Eismaschinen GmbH	Isernhagen, DE

2.2 Cell lines

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

2.3 Kits

Kit	Manufacturer
Total RNA Purification Kit	Jena Bioscience GmbH

2.4 Oligonucleotides

Target	Name	Sequence
CNN1	hCNN1.AS	5'-GTTGGCCTCAAAAATGTCGT-3'
	hCNN1.S	5'-AGGCTCCGTGAAGAAGATCA-3'
GAPDH	hGAPDH_II_Fw	5'-GGATTTGGTCGTATTGGG-3'
	hGAPDH_II_Rv	5'-GGAAGATGGTGATGGGATT-3'
MMP9	hMMP9_F	5'-CAGAGGTTGAACCCACAGT-3'
	hMMP9_R	5'-CCTCTGGCTTCGTCAGAATC-3'

Oligonucleotides were synthesized by Eurofins Genomics Germany GmbH.

2.5 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 (4 mM)	Applied Biosystems
DTT (0.1 M)	Invitrogen™ AG
Ethanol (99.9 %)	J.T.Baker®
FCCP	Sigma-Aldrich Co. LLC.
D-(+)-Glucose, 45 % solution	Sigma-Aldrich Co. LLC.
GlutaMAX™	Gibco® BRL
Hoechst 33342,	Invitrogen™ AG
IL-1 β	PeproTech®
M-MLV RT (200 U/ μ L)	Invitrogen™ AG

Continued on next page

(Continued)

Name	Manufacturer
NAC	Sigma-Aldrich Co. LLC.
NaHCO ₃	Carl Roth GmbH + Co. KG
NaOH, 1 N	Carl Roth GmbH + Co. KG
Oligomycin	Sigma-Aldrich Co. LLC.
PBS	Lonza Group AG
PDGF-BB	PeptoTech [®]
PowerUp [™] SYBR [™] GREEN Master Mix	Thermo Fisher Scientific Inc.
Roth Hexanukleotid Random-Primer-Mix	Carl Roth GmbH + Co. KG
RiboLock RNase Inhibitor (40 U/ μ L)	Thermo Fisher Scientific Inc.
Seahorse XF calibrant	Agilent Technologies, Inc.
Sodium Pyruvate (100 mM)	Gibco [®] BRL
TGF β	PeptoTech [®]

2.6 Media and supplements

Name	Manufacturer
FBS Gold Plus	Bio&SELL GmbH
M231	Gibco [®] BRL
Smooth Muscle Cell Growth Supplement	Thermo Fisher Scientific Inc.
XF Base Medium	Agilent Technologies, Inc.

2.7 Solutions

Name	Content
IL-1 β	10 μ g/mL IL-1 β 0.1 % BSA in PBS
NAC	0.25 M NAC in H ₂ O, ~pH 7 (adjusted)
PDGF-BB	100 μ g/mL PDGF-BB 0.1 % BSA in PBS

Continued on next page

(Continued)

Name	Content
TGF β	10 μ g/mL TGF β 0.1 % BSA in PBS

2.8 Consumables

Name	Manufacturer
384 Well Multiply PCR plates	Sarstedt AG & Co.
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 μ L	Sarstedt AG & Co.
Filter tip 200 μ L	Sarstedt AG & Co.
Filter tip 1000 μ L	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 μ L	Sarstedt AG & Co.
Pipette tip 200 μ L	Sarstedt AG & Co.
Pipette tip 1000 μ L	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.

Continued on next page

(Continued)

Name	Manufacturer
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.
Vial-Casette™	ChemoMetec A/S

2.9 Devices

Name	Manufacturer
7900HT Fast Real-Time PCR System	Thermo Fisher Scientific Inc.
BZ-X810 All-in-One Fluorescence Microscope	Keyence Corporation
BZ-X800 All-in-One Fluorescence Microscope POWER	Keyence Corporation
Centrifuge 5415 R	Eppendorf SE
Centrifuge 5702 R	Eppendorf SE
Clean Air®	Heraeus Holding GmbH
Eclipse TS100	Nikon Corporation
Eismaschine	ZIEGRA Eismaschinen GmbH
Hera Cell™	Heraeus Holding GmbH
Hera Cell™150	Heraeus Holding GmbH
HERAsafe®	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 Microprocessor pH Meter	HANNA Instruments
Pipet-X	Mettler-Toledo GmbH
Reax Top	Heidolph Instruments Labortechnik
Eppendorf™ Research™ pipettes (2.5 µL, 10 µL, 100 µL, 1000 µL)	Eppendorf SE

Continued on next page

(Continued)

Name	Manufacturer
Rotana 460 R	Andreas Hettich GmbH & Co. KG
SensoQuest labcycler	SensoQuest GmbH
XF24 Extracellular Flux Analyzer	Agilent Technologies, Inc.

2.10 Programs and Modules

Programs

Program	Version	Publisher
Affinity Designer	1.10	Serif (Europe) Ltd.
BZ-800 Analyzer	-	Keyence Corporation
BZ-800 Viewer	-	Keyence Corporation
Excel	Version 2205	Microsoft Corporation
GitHub	-	GitHub Inc
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
pylifter	0.4	github.com/konstantint/pylifter
python standard library	3.9	docs.python.org

Continued on next page

(Continued)

Module	Version	Information
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](#).

The scripts to generate this thesis and all shown analyses can be found at:
<https://github.com/ceryni5/Thesis>.

3

Methods

3.1 Cultivation and differentiation of HAoSMCs

HAoSMCs were used for the following experiments. Cells were kept at 37 °C and 5 % carbon dioxide (CO₂) whenever possible. For differentiation, cells were first treated with TGF β to induce a contractile phenotype. Afterward, HAoSMCs were stimulated with interleukin 1 beta (IL-1 β) 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 (6th 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 300 xg. 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:

160 μ L 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.

Table 3.1: Col I matrix composition

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

Differentiation 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 TGF β
- **Day 5:** The medium was replaced with 1 mL M231 + 1 % FBS + 10 ng/mL IL-1 β + 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 metalloproteinase (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:

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.

Table 3.2: Reaction mix for RT

component	concentration	volume (μL)
First Strand Buffer	5x	4
DTT	0.1 M	2
dNTP	4 mM	1
Hexanukleotid Random-Primer-Mix	30.6 nM	1
RiboLock RNase Inhibitor	40 U/ μl	1
M-MLV RT	200 U/ μl	1

qPCR

Table 3.3: Sample composition for qPCR

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

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

Table 3.4: qPCR cycle

step	time (s)	temperature ($^{\circ}\text{C}$)	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 C_q 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 C_q of the reference gene GAPDH was subtracted from the mean C_q of the gene of interest:

$$\Delta ct = ct(\text{gene of interest}) - ct(\text{GAPDH})$$

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

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

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 [analysis 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. (Agilent Technologies, Inc, 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 a Col I matrix!

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-CO₂ environment.

On the day of the assay, HAoSMCs were washed with 500 μ L PBS each. Afterward, the XF BASE medium was supplemented with 1 mM sodium pyruvate, 10 mM D-(+)-glucose, 2 mM GlutaMAX™, and 90 μ M NaOH. Cells were incubated for 1 h at 37 °C in a non-CO₂ environment with supplemented 500 μ L XF BASE medium. Toxins for disrupting 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 (μ M)	volume in cartridge (μ L)	concentration in well (μ M)
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 μ L Oligomycin
- 3 Repeats of:
 - Mixing (1 min)
 - Pause (2 min)
 - Detection of OCR and ECAR (4 min)
- Pause (2 min)
- Injection of 60 μ L FCCP
- 3 Repeats of:
 - Mixing (1 min)
 - Pause (2 min)
 - Detection of OCR and ECAR (4 min)
- Pause (2 min)
- Injection of 55 μ L 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 μ g/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 [analysis script](#).

3.4 CellROXTM assay

CellROXTM 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).

CellROXTM 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.

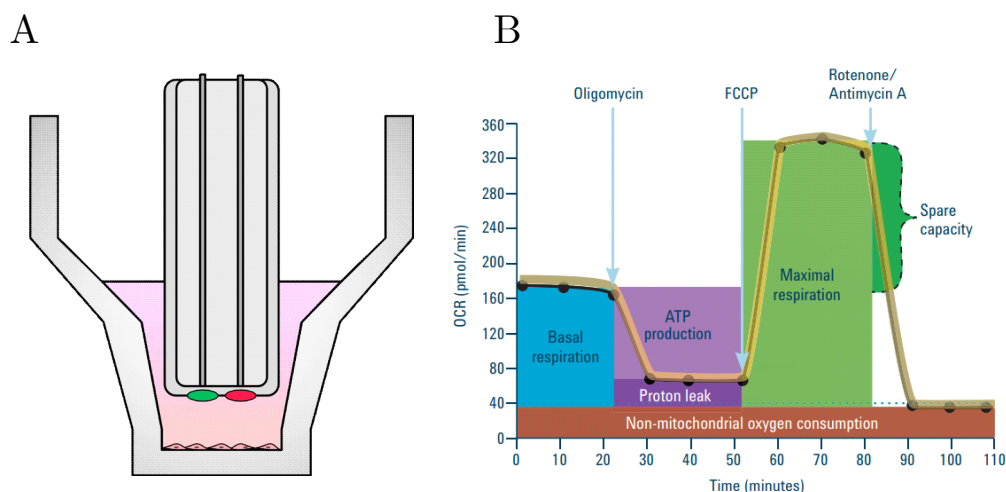


Figure 3.6: Basics of Seahorse Assay

(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). Figure adapted from (AgilentTechnologies, Inc, 2022). (B) Exemplary curve for OCR recorded over time and extractable properties of the respiratory chain (AgilentTechnologies, Inc., 2020).

CellROX™ Assay

For the assay, cells were washed with PBS. Then the boost was performed using variable concentrations of PDGF-BB in 300 μ L 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.

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

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

During the boost, cells were kept at 37°C in a 5% CO₂ 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 BZ-800 Analyzer.

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 [analysis scripts](#).

3.5 Curation of data for postGWAS analyses

Data for postGWAS analyses and co-visualization with the GWAS 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 CARDIoGRAMplusC4D summary statistics from Aragam et al. (2021) and the 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 used to generate 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 CARDIoGRAMplusC4D 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.

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

- **Ensembl Genome Annotations** The newest Ensembl build (Ensembl release 106) was downloaded via the [Ensembl FTP server](#). Features annotated as protein-coding

- (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 promoters
 - 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 \(UCSC\) 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 1257 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 data synchronization data between the underlying python environment and browser-side JS library. All in all, allowing real-time updates to the displayed data.

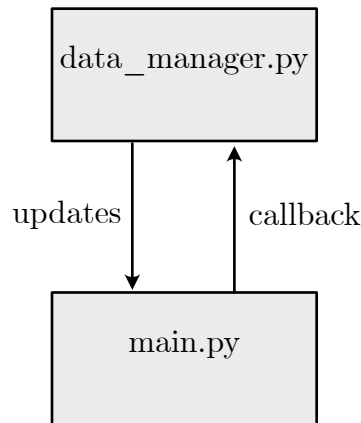


Figure 3.9: Architecture of the GWAS Navigator

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. For details please refer to the [source code of the GWAS Navigator](#).

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 the list of CAD-associated SNPs, the list of 241 proxy variants identified by the CARDIoGRAMplusC4D Consortium (Aragam et al., 2021), 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)

3 Methods

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 (Virtanen et al., 2020).

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

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

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

Additionally an enrichment factor (ef) was calculated as the hit rate for the biosample, divided by the background hit rate:

$$ef = \frac{\frac{nt}{n}}{\frac{mt}{m}} = \frac{nt \times m}{n \times mt}$$

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](#).

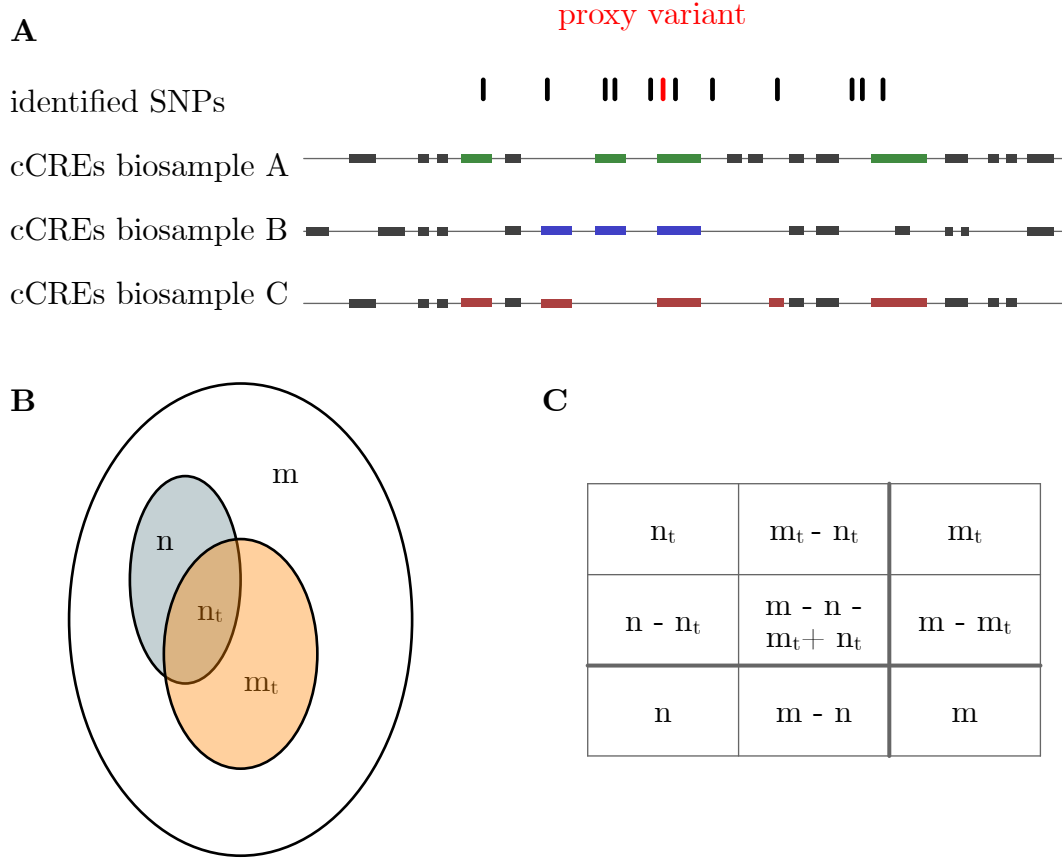


Figure 3.10: Enrichment analysis for cCREs overlapping with CAD risk SNPs

(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); (m_t) 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 Differentiation

To characterize the influence of PDGF-BB stimulation on HAoSMCs, the cells were first treated with TGF β 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-1 β 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 levels were normalized with the expression of the housekeeping gene GAPDH.

Figure 4.1 (top panel) illustrates that stimulation of HAoSMCs cultivated on a Col I matrix with TGF β causes a significant increase in CNN1 expression (+- vs. - -). After further stimulation with PDGF-BB and IL-1 β , 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 noticeable for HAoSMCs cultivated on plastic. However, this effect did not reach significance after four biological repeats. Additionally, stimulation of HAoSMCs on plastic with TGF β , followed by stimulation with PDGF-BB and IL-1 β , yields a significantly lower expression of CNN1 (++ Matrix vs. ++ Plastic).

As seen in the bottom panel of figure 4.1, no statistically significant trends can be observed after four biological repeats 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 noticeable difference being between HAoSMCs treated first with TGF β as well as with PDGF-BB and IL-1 β (++ Matrix vs. ++ Plastic, $p=0.063$).

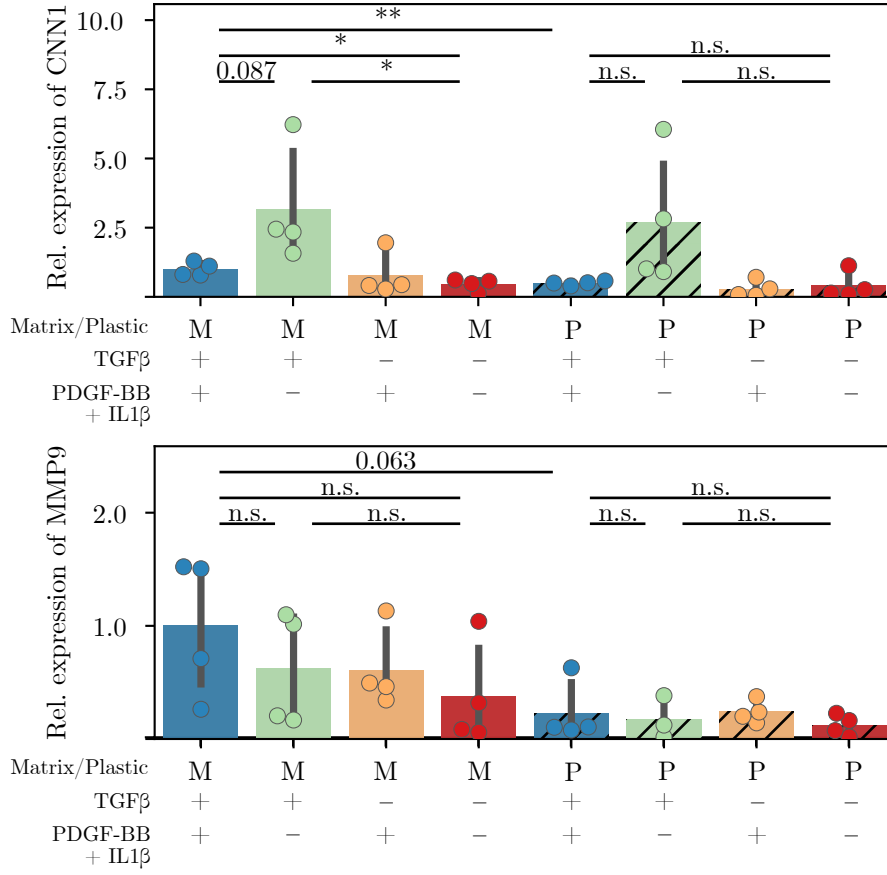


Figure 4.1: Relative Expression of CNN1 and MMP9 in HAoSMCs
qPCR analysis for expression of contractile marker CNN1 (top) and synthetic marker MMP9 (bottom) in HAoSMCs differentiated with different combinations of cytokines: ++: four days with TGFβ followed by two days with IL-1β and PDGF-BB; + -: four 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: * : $p < 0.05$; ** : $p < 0.01$

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 because the Col I matrix does not fit into the confined compartment created by the piston to record the 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.

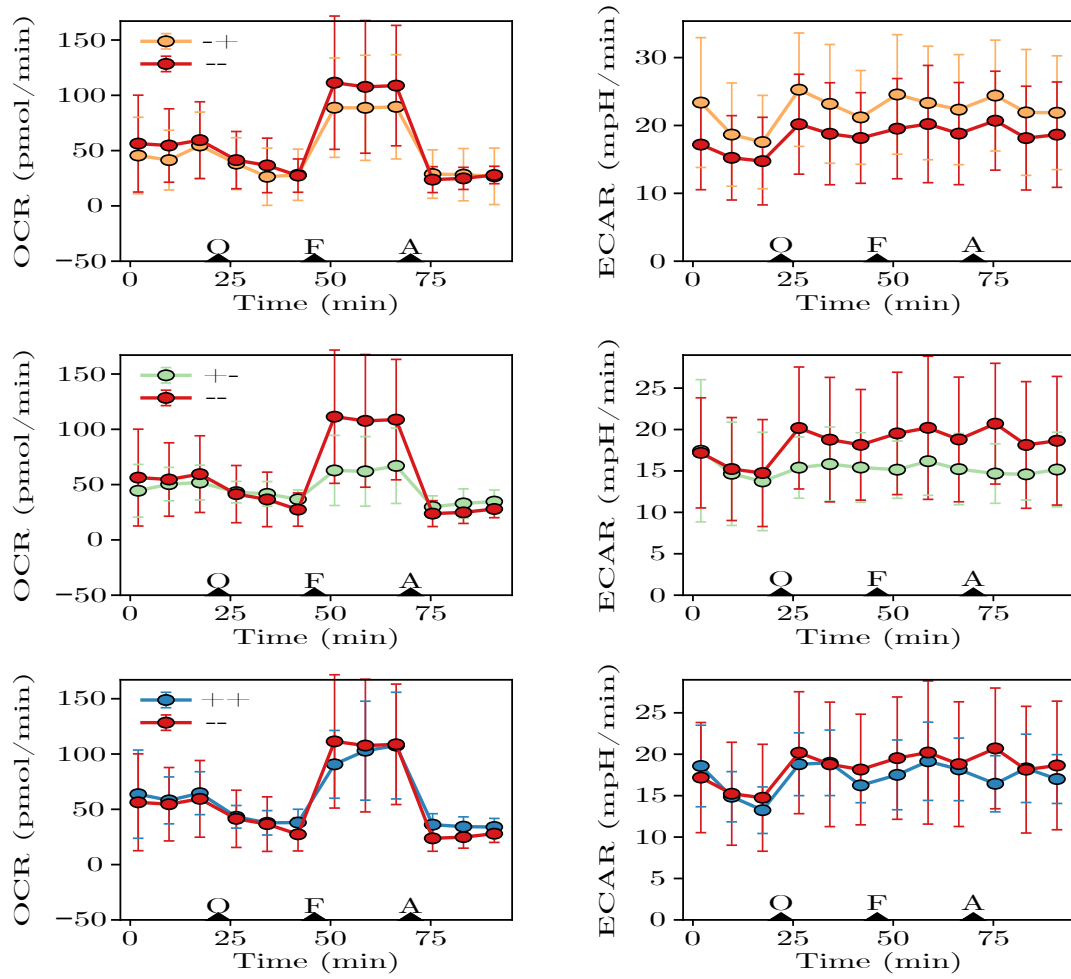


Figure 4.2: OCR and ECAR of HAoSMCs

Seahorse assay for HAoSMCs differentiated with different combinations of cytokines. ++: four days with TGF β followed by two days with IL-1 β and PDGF-BB; +-: four days with TGF β followed by two days without stimulation; -+: four days without stimulation followed by two days with IL-1 β and PDGF-BB; —: six days without stimulation. OCR and ECAR are shown for +- (top), ++ (middle) and ++ (bottom) compared 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.

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 the addition of toxins impacting the respiratory chain (compare to figure 3.6 B). After inhibition of the adenosine triphosphate (ATP) synthase with Oligomycin, the basal OCR drops, revealing the proportion of the OCR required for 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 adding Oligomycin, most likely because the cells compensate for the loss of mitochondrial ATP production via increased glycolysis.

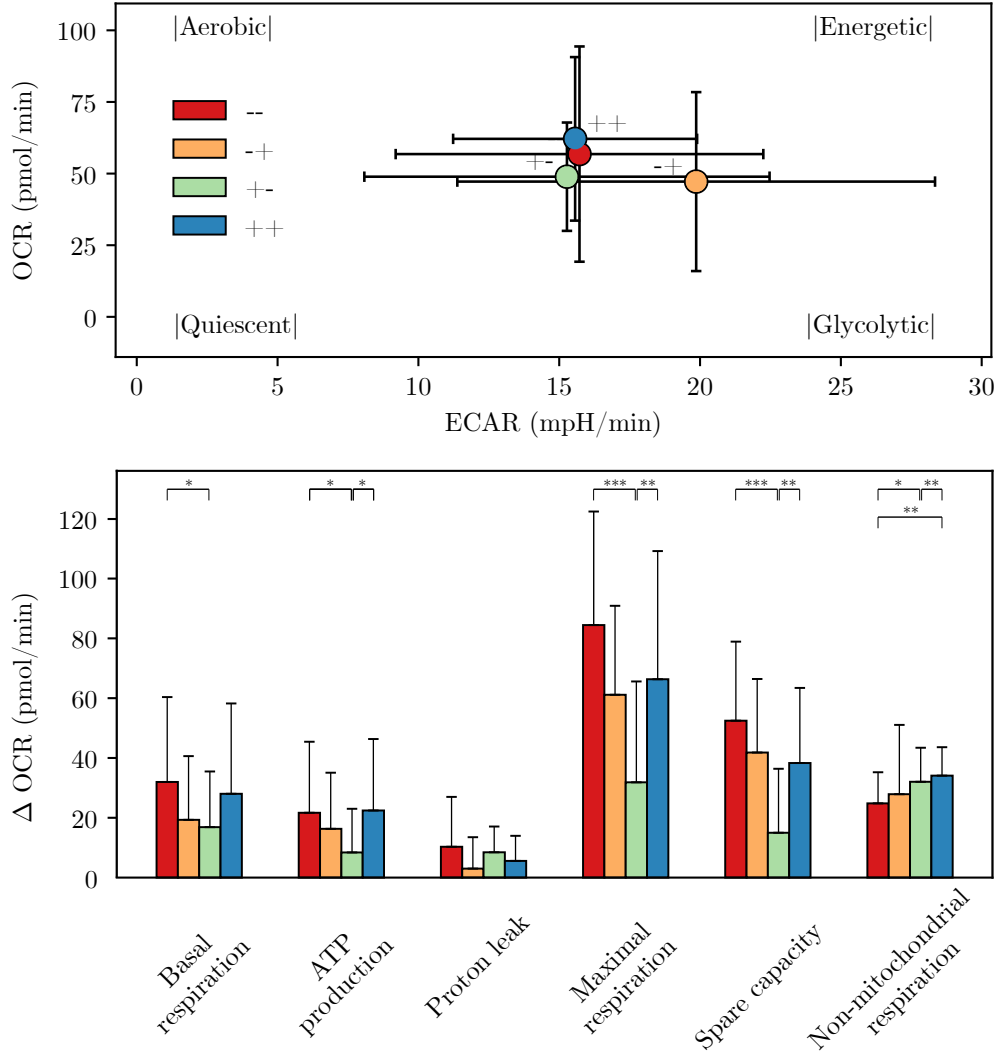


Figure 4.3: Energy profile of HAoSMCs

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 Welch'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-1 β and PDGF-BB (-+) (fig. 4.3, A). More pronounced differences can be observed when examining the characteristics of the respiratory chain. Stimulation with only TGF β causes a significant decrease in basal respiration, ATP production, maximal respiration, and spare capacity (figure 4.3 B). Further stimulation with IL-1 β and PDGF-BB then causes a significant increase of these parameters to similar levels as in initially dedifferentiated HAoSMCs (figure 4.3 B).

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-BB boost of dedifferentiated HAoSMCs

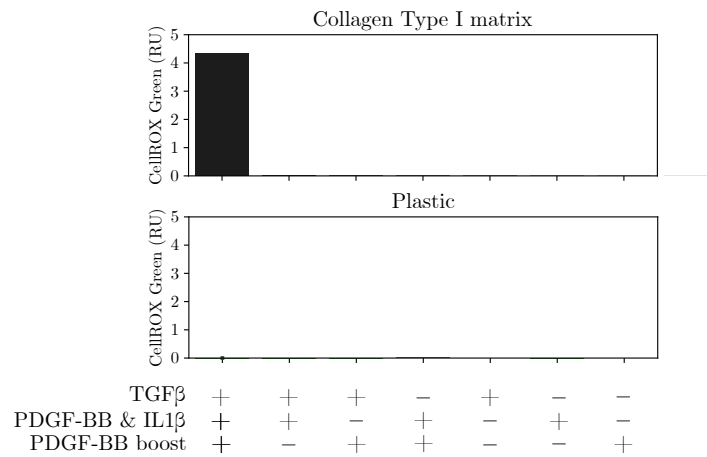


Figure 4.4: Boost with PDGF-BB induces the generation of ROS.

CellROXTM assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGF β ; followed by two days with IL-1 β 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 shown signal was calculated according to section 3.4 as the CellROXTM 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 TGF β , followed by two days with IL-1 β and PDGF-BB, followed by a 2 h boost with PDGF-BB, triggered noticeable ROS generation 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

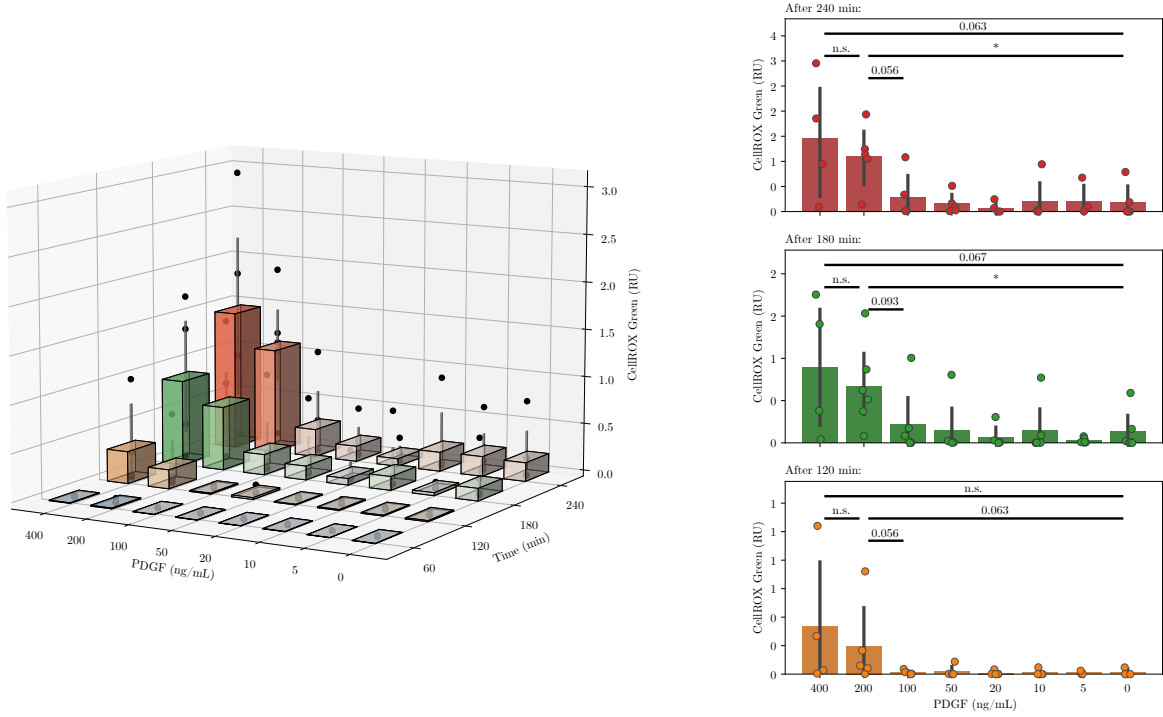


Figure 4.5: PDGF-BB boost titration

CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGF β ; followed by two days with IL-1 β 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: * : $p < 0.05$.

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 TGF β as well as two days with 10 ng/mL IL-1 β and 10 ng/mL PDGF-BB, were boosted with different concentrations of PDGF-BB (0 - 400 ng/mL). Signal was detected after 60, 120, 180, and 240 min in HBSS (figure S.1). As seen in figure 4.5, the 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 compared 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 reproducible. On the one hand, the signal was extremely high. On the other hand, the other two repeats, it collapsed.

4 Results

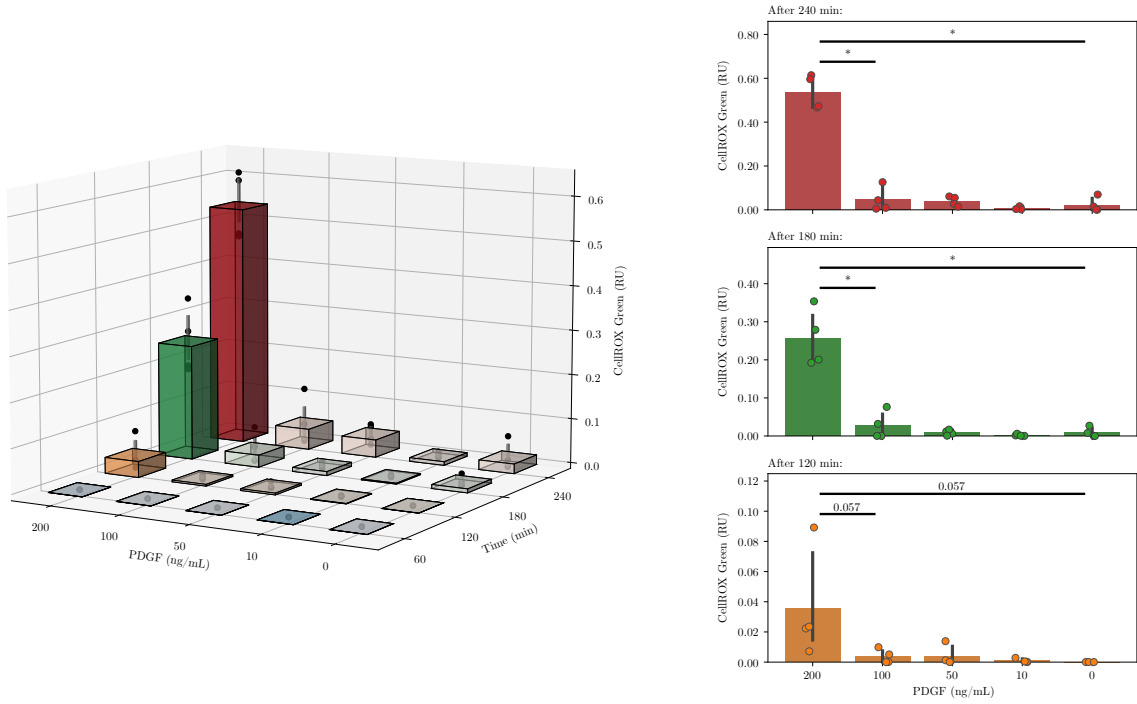


Figure 4.6: PDGF-BB boost titration - normalized

CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGF β ; followed by two days with IL-1 β and PDGF-BB; followed 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 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. 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: * : $p < 0.05$. Note that not every biological repeat covered *all* PDGF-BB concentration.

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

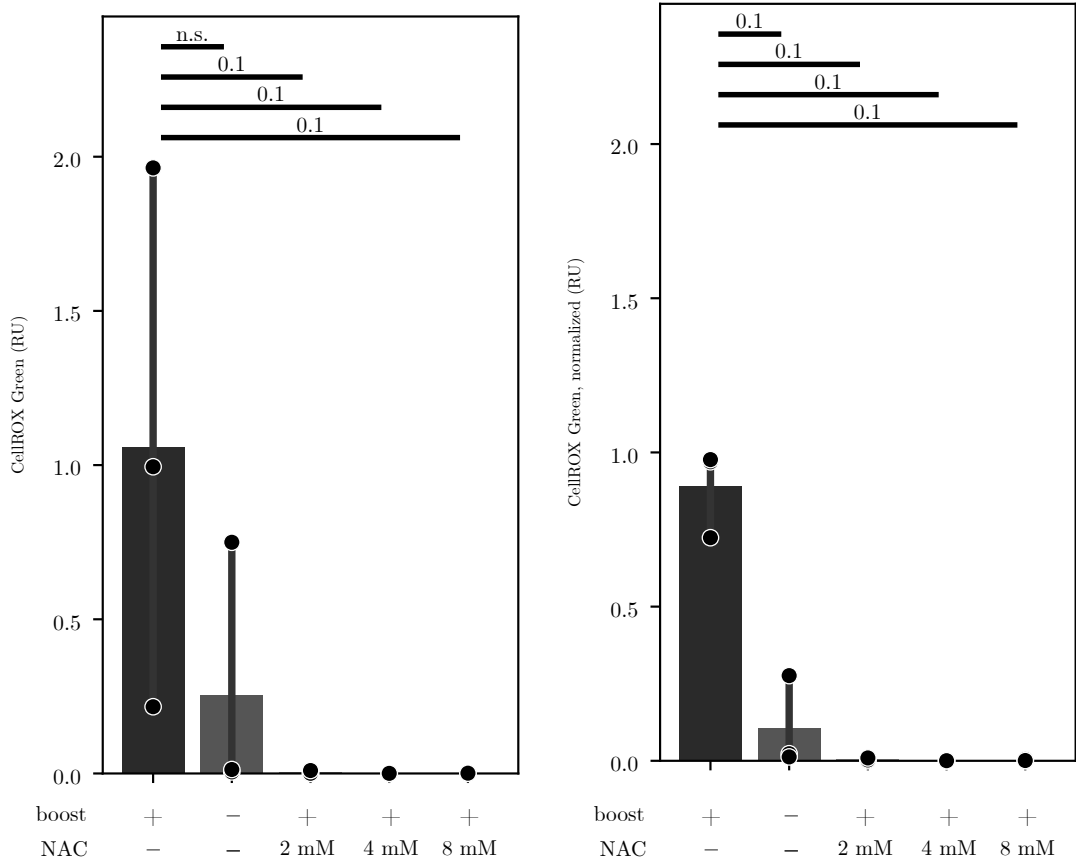


Figure 4.7: ROS generation due to PDGF-BB boost can be rescued with NAC

CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: four days with TGF β ; followed by two days with IL-1 β 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 = 3) biological repeats was performed using Mann-Whitney U test: * : $p < 0.05$. pos Ctrl: not treated with NAC, negCtrl: no boost with 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. ROS generation was quenched by adding 2, 4, or 8 mM of NAC. Indeed, a clear trend can be observed: HAoSMCs treated with NAC showed 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 the 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% CO₂ 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 Navigator

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 index exists, please consult 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_CCRC	4,451,476
	ENCODE_CCRC_META	107
total	-	6,284,781 (\approx 25.75 GB)

4 Results

The first step toward visualization of GWAS summary statistics was curating 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 database’s entity-relationship (ER) diagram in figure 4.9.

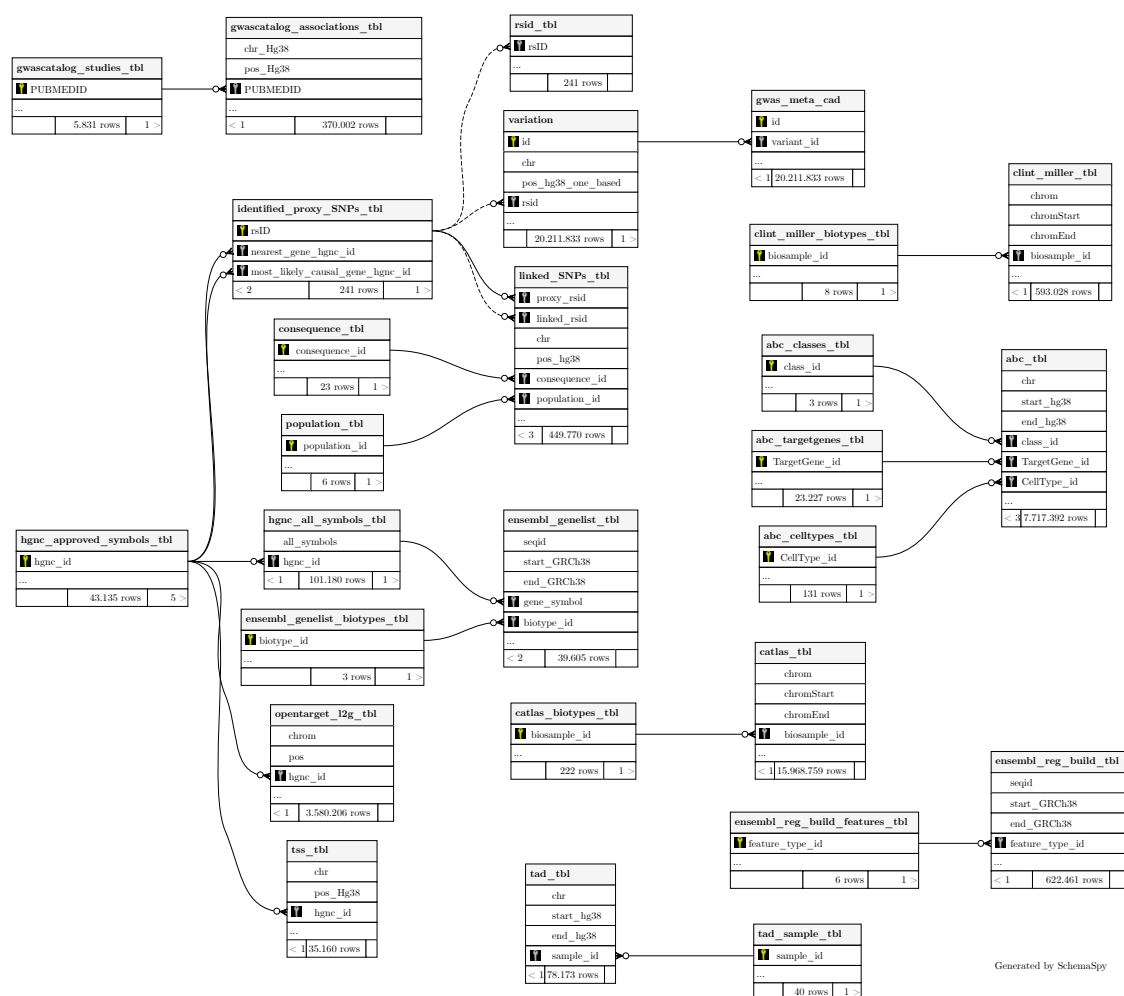


Figure 4.9: Entity-Relationship diagram of the database

Fields and relationships of the tables listed in table 4.8. On spelled out columns an index exists or they are primary or foreign keys. The diagram was generated via SchemaSpy.

4 Results



A tool to visualize GWAS summary statistics for intuitive exploration.

GWAS data for coronary artery disease. Annotated with associated phenotypes from GWAS catalog, genomic features from ensembl.org, aligned with scATAC-seq data from CATlas as well as promotor enhancer interactions in form of ABC scores.

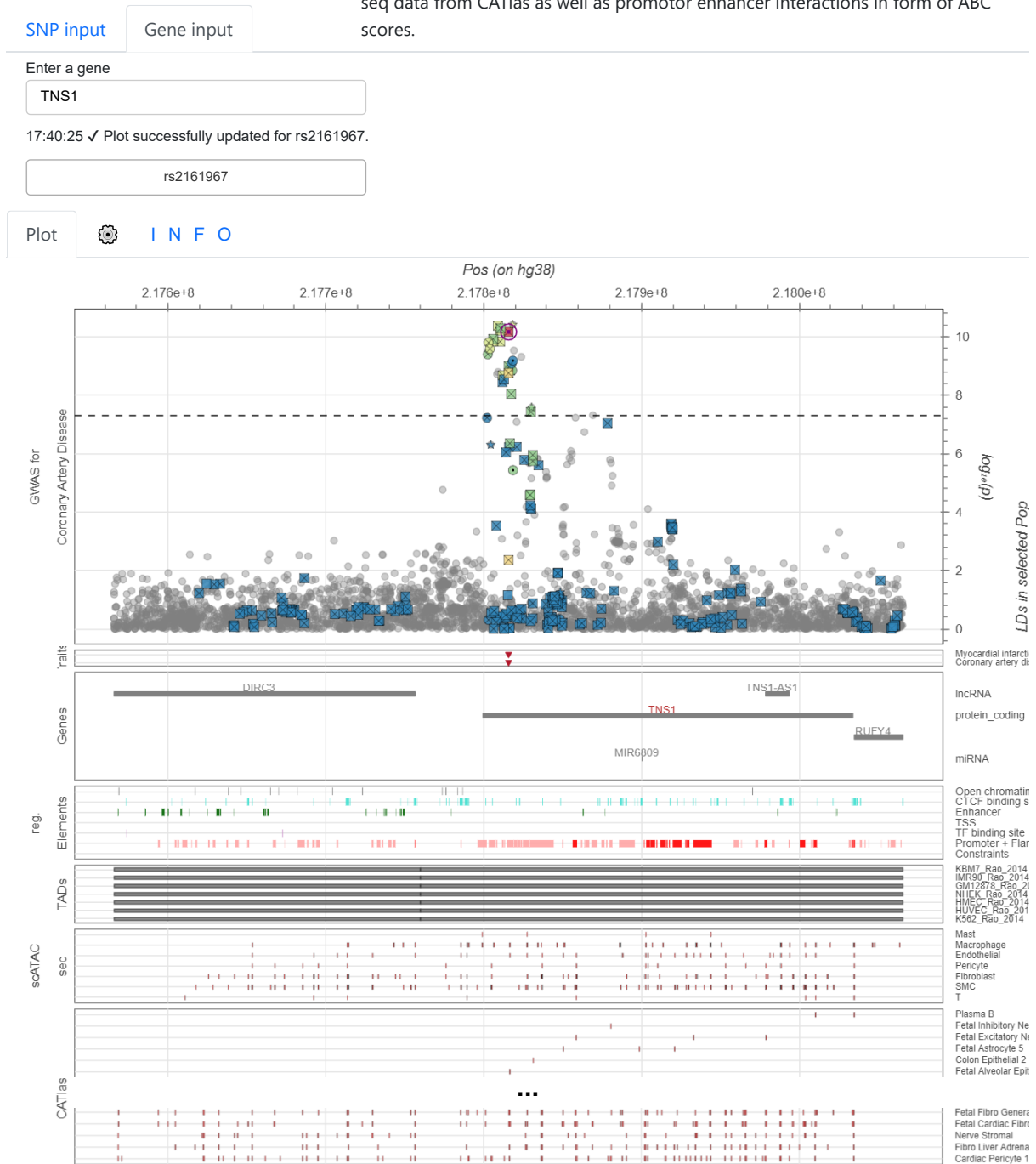


Figure 4.10: The GWAS Navigator

General content of the GWAS Navigator. The tool contains a manhattan plot with GWAS 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 GWAS summary statistics was built according to section 3.6. As shown in figure 4.10, the GWAS 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 is chosen, the tool displays the GWAS summary statistics in a 500 kb window centered around the selected variant to the output panel. GWAS summary statistics are visualized as a zoomed-in Manhattan 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 GWAS 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 > 0.3) 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 with no elements in the visualized region.

The GWAS Navigator also has a settings tab where individual tracks can be hidden.

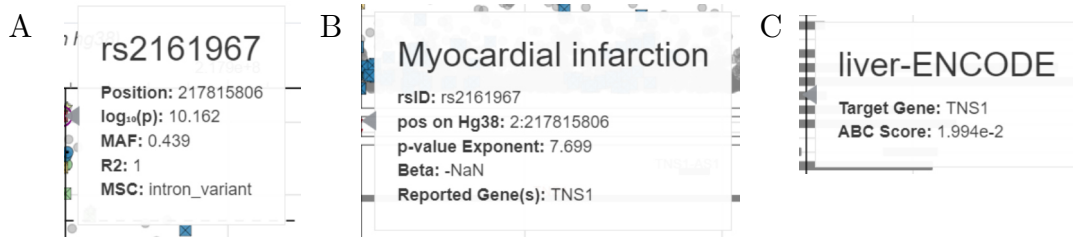


Figure 4.11: The GWAS Navigator - hover effect

Exemplary hover effects for features displayed in the GWAS 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 not displayed in the plot are ENCODE cCREs subjected to an enrichment analysis. The annotated biosamples are checked for significant enrichment of cCREs that overlap with proxy SNPs identified by the ARDIoGRAMplusC4D Consortium (Aragam et al., 2021) or variants in LD with these ($r^2 > 0.6$). For more details, please refer to section 3.7.

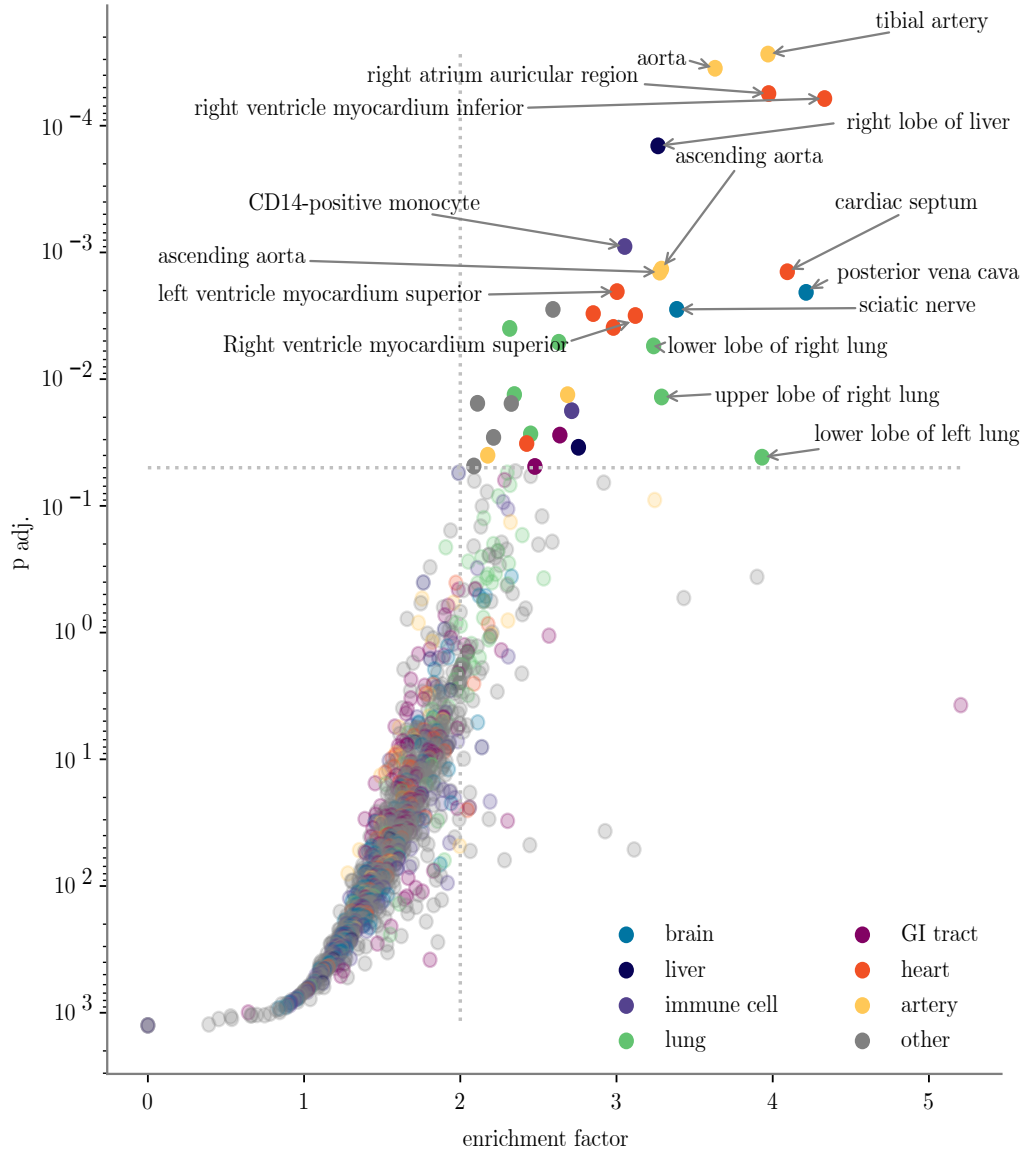


Figure 4.12: Enrichment analysis for overlap of CAD GWAS proxy variants with tissue specific cCREs

p-values and enrichment factors for the overlap of CARDIoGRAMplusC4D proxy variants (and variants in LD) with tissue specific cCREs. For details please refer to section 3.7.

As seen in figure 3.10, statistically significant enrichment ($p_{adj.} < 0.05$) was observed for 34 biosamples (table 4.14). 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 statistically significant overlap between CARDIoGRAMplusC4D Consortium 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

Table 4.14: List of all biosamples that show significant enrichment

nt	mt	n	m	p	p_adj	ef	biosample_term_name	cellosaurus_attr_2	cellosaurus_attr_3
23	60010	165	1709065	2.16e-08	2.72e-05	3.96	tibial artery	artery	artery
25	71318	165	1709065	2.80e-08	3.52e-05	3.63	aorta	aorta	artery
22	57342	165	1709065	4.43e-08	5.57e-05	3.97	right atrium auricular region	cardiac atrium	heart
20	47820	165	1709065	4.85e-08	6.10e-05	4.33	Right ventricle myocardium inferior	myocard	heart
26	82463	165	1709065	1.14e-07	1.44e-04	3.26	right lobe of liver	liver	liver
25	84828	165	1709065	7.14e-07	8.97e-03	3.05	CD14-positive monocyte	monocytes	immune cell
22	69308	165	1709065	1.07e-06	1.35e-03	3.28	ascending aorta	aorta	artery
17	43017	165	1709065	1.13e-06	1.42e-03	4.09	cardiac septum	cardiac septum	heart
22	69542	165	1709065	1.13e-06	1.43e-03	3.27	ascending aorta	aorta	artery
24	82774	165	1709065	1.62e-06	2.03e-03	3.00	left ventricle myocardium superior	myocard	heart
16	39330	165	1709065	1.64e-06	2.06e-03	4.21	posterior vena cava	neuronal cell	brain
20	61185	165	1709065	2.23e-06	2.81e-03	3.38	sciatic nerve	neuronal cell	brain
29	115843	165	1709065	2.24e-06	2.81e-03	2.59	dermis microvascular lymphatic vessel endothelial cell	endothelial cell	lymphatic vessel
25	90831	165	1709065	2.41e-06	3.03e-03	2.85	heart left ventricle	heart left ventricle	heart
22	73021	165	1709065	2.50e-06	3.14e-03	3.12	Right ventricle myocardium superior	myocard	heart
23	79943	165	1709065	3.10e-06	3.90e-03	2.98	left ventricle myocardium inferior	myocard	heart
34	152031	165	1709065	3.16e-06	3.97e-03	2.31	lung microvascular endothelial cell	endothelial cell	lung
27	106321	165	1709065	4.07e-06	5.11e-03	2.63	right lung	lung	lung
20	63984	165	1709065	4.34e-06	5.46e-03	3.23	lower lobe of right lung	lung	lung
30	132446	165	1709065	1.05e-05	1.32e-02	2.34	lung microvascular endothelial cell	endothelial cell	lung
24	92520	165	1709065	1.05e-05	1.32e-02	2.68	dermis blood vessel endothelial cell	endothelial cell	artery
18	56694	165	1709065	1.09e-05	1.38e-02	3.28	upper lobe of right lung	lung	lung
36	176705	165	1709065	1.23e-05	1.55e-02	2.11	subcutaneous abdominal adipose tissue	adipocyte	adipose tissue
30	133565	165	1709065	1.23e-05	1.55e-02	2.32	ovary	ovary	ovary
23	87812	165	1709065	1.41e-05	1.77e-02	2.71	CD14-positive monocyte	monocytes	immune cell
26	109915	165	1709065	2.15e-05	2.70e-02	2.45	left lung	lung	lung
23	90330	165	1709065	2.19e-05	2.76e-02	2.63	left colon	colon	GI tract
31	145102	165	1709065	2.29e-05	2.88e-02	2.21	endothelial cell of umbilical vein	endothelial cell	embryo
26	111050	165	1709065	2.56e-05	3.21e-02	2.42	heart	heart	heart
21	78914	165	1709065	2.74e-05	3.44e-02	2.75	left lobe of liver	liver	liver
31	147590	165	1709065	3.17e-05	3.98e-02	2.17	dermis blood vessel endothelial cell	endothelial cell	artery
13	34248	165	1709065	3.28e-05	4.12e-02	3.93	lower lobe of left lung	lung	lung
33	163804	165	1709065	3.85e-05	4.84e-02	2.08	kidney	kidney	kidney
24	100336	165	1709065	3.88e-05	4.88e-02	2.47	mucosa of gallbladder	gallbladder	GI tract

5

Discussion

5.1 PDGF-BB signaling induces a synthetic phenotype in HAoSMCs

The crucial role of vSMCs in atherogenesis has been the subject of extensive research for the last few decades (Grootaert and Bennett, 2021; Yap et al., 2021). Traditionally it has been assumed that vSMCs adopt a protective role by stabilizing the atherosclerotic 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 the dedifferentiation process is an initial mesenchymal-like phenotype. This phenotype exhibits a proliferative phenotype and reduced expression of contractile markers (Yap et al., 2021). This phenotype is thought to be initiated by the TF KLF4, which induces the 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 processes disrupt myocardin/serum response factor (SRF) mediated expression of contractile genes. The mesenchymal-like phenotype is postulated as a precursor for other dedifferentiated vSMCs phenotypes (Yap et al., 2021).

The contractile differentiated vSMCs phenotype is constantly maintained by myocardin/SRF signaling (Long et al., 2008), as well as external stimulation by the extracellular matrix (ECM) and cytokines such as TGF β (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 with TGF β for four days, 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). This trend is possibly an adaptation to the energetic needs of the contractile phenotype, which is considered quiescent (Dobnikar et al., 2018). Further stimulation with PDGF-BB and IL-1 β for two days yields a the decrease expression of CNN1, on the verge of significance ($p=0.087$). Moreover, the energy metabolism changes again, characterized by a rebound of basal mitochondrial respiration, ATP production, and maximal respiration to similar levels as initially dedifferentiated VSMCs (figure 4.3 B).

Another important aspect of phenotypic transition and plaque development is the remodeling of the ECM by MMPs (Johnson, 2017). Our experiments hint toward a possible increase of MMP9 expression for HAoSMCs cultivated on Col I matrix (figure 4.1 bottom) after PDGF-BB-induced dedifferentiation. However, this trend remained not significant in four biological replicates ($p = 0.063$). MMP9 is an important component of atherosclerosis (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) underlines the bi-directionality of the ECM-vSMC-interactions and the complexity of vSMC dedifferentiation.

Of course, the PDGF-BB-induced phenotype can not be grasped with only two markers and requires a more in-depth analysis, for example via RNA-sequencing.

5.2 CellROXTM Green is suitable for assessing ROS generation in HAoSMCs

Evaluating the response to further stimulation with PDGF-BB, the CellROXTM Assay confirmed a result previously observed in the group (unpublished). Stimulation of HAoSMCs cultivated on Col I matrix and treated for four days with TGF β and two days with PDGF-BB and IL-1 β 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). In addition, it was observed that the signal intensity highly depends on the incubation time. While the trend for each biological repeat is clear, the variance between repeats is in a similar range. The assay is working reliably but could greatly benefit from retroactive normalization (figure 4.6) or further optimization of reproducibility - reducing the required amount of biological repeats. An alternative option for normalization could be provided by direct stimulation with H₂O₂ to be used as a reference. Finally, the robustness of the assay may be improved by the exploration of different CellROXTM Green concentrations.

Finally, a recovery experiment was performed. Before and during the boost, cells were co-incubated with NAC. NAC is a popular and potent antioxidant used in cell culture experiments, and it likely acts by being metabolized into sulfane sulfur species that scavenge ROS in the mitochondria (Ezerina et al., 2018). Cells treated with NAC showed only little to no CellROXTM Green signal. Even though this trend remained not significant after three replicates, it supports 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 ($\cong 8.25$ nM) is physiologically relevant. Unfortunately, cytokine concentrations are usually assessed as plasma concentrations, and no *in vivo* data exists for local concentrations during paracrine signaling. While the manufacturer describes the half maximal effective concentration (EC₅₀) 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) up to 10 nM ($\cong 240$ ng/mL) PDGF-BB when assessing the dose-response relationship between cAMP formation after PDGF-BB stimulation of SMCs. Newman et

al. (2021b) used 50 ng/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 PDGFR β . 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 the transcription factor STAT1 (STAT1), which upon 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, STAT1 can be imported into mitochondria, where it interacts with respiratory complexes and triggers the generation of ROS (Wang et al., 2018a) during hepatic apoptosis (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 Cell-ROXTM Green signal would occasionally only develop outside the controlled environment in the incubator. This suggests that the PDGF-BB is not the sole trigger of ROS generation. Repeating the experiment under better-controlled conditions would be a great option to follow up on this idea.

We additionally tried to assess oxidative stress with an anti-8-oxoguanine antibody that detects 8-oxoguanine, a base modification often observed in the presence of ROS (Leon et al., 2016). This attempt was not successful because the starvation of HAoSMCs for seven days in M231 supplemented with 1 % FBS was sufficient to induce oxidative damage to the genome (figure S.2).

5.3 The GWAS Navigator

Like all primates, humans are extremely visual creatures. We have evolved specialized brain structures for processing visual stimuli (Kaas and Balaram, 2014), granting us superior recognition of visual patterns (Mattson, 2014). Thus, making data visualization tools powerful and important resources for interactive data exploration and scientific communication. The GWAS Navigator was developed to display CARDIoGRAMplusC4D Consortium (Aragam et al., 2021) summary statistics in a comprehensive and visually appealing format for medical researchers. In an iterative process, many possible implementation approaches were explored, finally resulting in 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 data collection at the heart of data science. Databases provide many advantages over data stored in spreadsheets, such as access speed, maintainability, and multiuser access. They are designed to hold large data collections and provide secure and fast access by querying via specifically designed database engines. Relational databases, like SQLite, are the most popular way for flexibly representing data in tables comprising columns and rows. They are usually queried and manipulated with commands in 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. The package combines the elegant visualization resources for rendering data with hypertext markup language (HTML), CSS, and JS to the browser with the powerful data processing capabilities of python. All are bundled into one easy-to-learn ecosystem, providing a level of abstraction required for a prototype’s construction. Additionally, the bokeh server makes the application easily deployable for potential use on a local network (Bokeh Development Team, 2022).

To summarize, the GWAS Navigator grants an overview of 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 group. It provides basic functionality and the possibility to implement many additional features. Options range from basic improvements to usability in the formed tissue-specific annotations to the displayed tracks and the selection tool to the adoption of new datasets.

5.4 Overlap of CAD associated variants with regulatory elements is enriched in heart, artery, and lung tissue

In addition to providing the basis for the GWAS Navigator, the database also makes the data easily accessible for follow-up studies. The curated data are utilized in an initial post-GWAS analysis, scanning for CARDIoGRAMplusC4D Consortium (Aragam et al., 2021) proxy variant enrichment in biosample-specific cCREs via Fisher’s exact test. This way identifying 34 biosamples (of 1257 tested) that show significant overrepresentation (figure 4.12 and table 4.13). After annotating these biosamples, over 40 % (14/34) of enriched biosamples stem from heart or artery tissue and are directly affected by atherosclerosis. An additional 20 % was annotated to stem from lung tissue, an observation in line with the frequently reported association between heart- and lung diseases (Carter et al., 2019; Han et al., 2007). The association of heart- and lung diseases prevails even after adjustment for shared risk factors such as tobacco usage or age. Additionally, Au Yeung et al. (2018) demonstrated via Mendelian randomization that greater forced expiratory volume in 1 second (FEV_1) decreases the risk of CAD. Still, the causality of this relationship remains unclear. While it is tempting to speculate that impaired lung function or systematic inflammation by chronic obstructive pulmonary disease (COPD) results in an elevated risk for cardiovascular diseases, such hypotheses are difficult to evaluate due to reverse causation (Nowak, 2018). CAD might also be a risk factor for lung diseases, or both pathologies could share additional mutually relevant confounding factors. Similarly, the identification of lung tissue in our analysis might point toward the contribution of the lung during the development of CAD or a shared genomic predisposition of heart- and lung disease. Following up on systemic inflammation, the immune cells in which cCREs enrich are $CD14^+$ monocytes (table 4.14), a cell type known for the secretion of proinflammatory cytokines during injury or inflammation (Kapellos et al., 2019). Interestingly, $CD14^{++}CD16^+CCR2^+$ and $CD14^{++}CD16^-CCR2^+$ 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. Consequently, based on the enhancer promotor connection, one may be able to identify potentially affected genes.

Bibliography

WHO (2022). *Cardiovascular Diseases*. URL: <https://www.who.int/health-topics/cardiovascular-diseases> (visited on 07/21/2022).

Doran, Amanda C., Meller, Nahum, and McNamara, Coleen A. (2008). Role of Smooth Muscle Cells in the Initiation and Early Progression of Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 28, 812–819. DOI: [10.1161/ATVBAHA.107.159327](https://doi.org/10.1161/ATVBAHA.107.159327).

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](https://doi.org/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](https://doi.org/10.1093/cvr/cvab046).

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](https://doi.org/10.1161/ATVBAHA.121.316600).

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

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](https://doi.org/10.1016/j.jacc.2018.11.063).

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

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

Centers for Disease Control and Prevention (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. *NCHS Data Brief*, 8.

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

Bibliography

- 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/ehs296](https://doi.org/10.1093/eurheartj/ehs296).
- Tucker, William D., Arora, Yingyot, and Mahajan, Kunal (2022). 1. 1, Treasure Island (FL): StatPearls Publishing.
- Tabas, Ira, García-Cardena, Guillermo, and Owens, Gary K. (2015). Recent Insights into the Cellular Biology of Atherosclerosis. *The Journal of Cell Biology* *209*, 13–22. DOI: [10.1083/jcb.201412052](https://doi.org/10.1083/jcb.201412052).
- Goumans, Marie-José and Dijke, Peter ten (2018). TGF-beta Signaling in Control of Cardiovascular Function. *Cold Spring Harbor Perspectives in Biology* *10*, a022210. DOI: [10.1101/cshperspect.a022210](https://doi.org/10.1101/cshperspect.a022210).
- Batlle, Eduard and Massagué, Joan (2019). Transforming Growth Factor-beta Signaling in Immunity and Cancer. *Immunity* *50*, 924–940. DOI: [10.1016/j.immuni.2019.03.024](https://doi.org/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-beta and Bone Morphogenetic Protein 4. *The Journal of Biological Chemistry* *286*, 28097–28110. DOI: [10.1074/jbc.M111.236950](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1111/1440-1681.12478).
- Andrae, Johanna, Gallini, Radosa, and Betsholtz, Christer (2008). Role of Platelet-Derived Growth Factors in Physiology and Medicine. *Genes & Development* *22*, 1276–1312. DOI: [10.1101/gad.1653708](https://doi.org/10.1101/gad.1653708).

Bibliography

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.cytogfr.2004.03.004).
- He, Chaoyong et al. (2015). PDGFRbeta Signalling Regulates Local Inflammation and Synergizes with Hypercholesterolaemia to Promote Atherosclerosis. *Nature Communications* *6* (1), 7770. DOI: [10.1038/ncomms8770](https://doi.org/10.1038/ncomms8770).
- Newman, Alexandra A. C. et al. (2021a). Multiple Cell Types Contribute to the Atherosclerotic Lesion Fibrous Cap by PDGFRbeta and Bioenergetic Mechanisms. *Nature Metabolism* *3* (2), 166–181. DOI: [10.1038/s42255-020-00338-8](https://doi.org/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](https://doi.org/10.1038/s41580-020-0230-3).
- Zeida, Ari et al. (2019). Catalysis of Peroxide Reduction by Fast Reacting Protein Thiols. *Chemical Reviews* *119*, 10829–10855. DOI: [10.1021/acs.chemrev.9b00371](https://doi.org/10.1021/acs.chemrev.9b00371).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.chembiol.2014.02.020).
- Flint, Jonathan (2013). GWAS. *Current Biology* *23*, R265–R266. DOI: [10.1016/j.cub.2013.01.040](https://doi.org/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](https://doi.org/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](https://doi.org/10.1038/s41467-020-20188-y).

Bibliography

- Holdt, Lesca M. and Teupser, Daniel (2012). Recent Studies of the Human Chromosome 9p21 Locus, Which Is Associated with Atherosclerosis in Human Populations. *Arteriosclerosis, Thrombosis, and Vascular Biology* *32*, 196–206. DOI: [10.1161/ATVBAHA.111.232678](https://doi.org/10.1161/ATVBAHA.111.232678).
- Erdmann, Jeanette et al. (2018). A Decade of Genome-Wide Association Studies for Coronary Artery Disease: The Challenges Ahead. *Cardiovascular Research* *114*, 1241–1257. DOI: [10.1093/cvr/cvy084](https://doi.org/10.1093/cvr/cvy084).
- Khera, Amit V. and Kathiresan, Sekar (2017). Genetics of Coronary Artery Disease: Discovery, Biology and Clinical Translation. *Nature Reviews. Genetics* *18*, 331–344. DOI: [10.1038/nrg.2016.160](https://doi.org/10.1038/nrg.2016.160).
- 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](https://doi.org/10.1101/2021.05.24.21257377).
- Slatkin, Montgomery (2008). Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. *Nature reviews. Genetics* *9*, 477–485. DOI: [10.1038/nrg2361](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1038/s41586-020-2493-4).
- 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](https://doi.org/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](https://doi.org/10.1038/s41586-021-03446-x).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1038/nature14590).

Bibliography

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3389/fcvm.2019.00089).
- 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](https://doi.org/10.1007/s00769-011-0769-y).
- AgilentTechnologies, Inc (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).
- AgilentTechnologies, Inc. (2020). *AGILENT Seahorse XF Cell Mito Stress Test Kit*.
- 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 PDGFRbeta and Bioenergetic Mechanisms. *Nature Metabolism* *3* (2), 166–181. DOI: [10.1038/s42255-020-00338-8](https://doi.org/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](https://doi.org/10.1186/1479-7364-4-3-202).

Bibliography

- Virtanen, Pauli et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17, 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1172/JCI86924).
- Ezerina, Daria et al. (2018). N-Acetyl Cysteine Functions as a Fast-Acting Antioxidant by Triggering Intracellular H2S and Sulfane Sulfur Production. *Cell Chemical Biology* 25, 447–459.e4. DOI: [10.1016/j.chembiol.2018.01.011](https://doi.org/10.1016/j.chembiol.2018.01.011).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/j.jmb.2007.03.072).
- Leon, Julio et al. (2016). 8-Oxoguanine Accumulation in Mitochondrial DNA Causes Mitochondrial Dysfunction and Impairs Neurogenesis in Cultured Adult Mouse Cortical Neurons under Oxidative Conditions. *Scientific Reports* 6 (1), 22086. DOI: [10.1038/srep22086](https://doi.org/10.1038/srep22086).
- 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](https://doi.org/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](https://doi.org/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).
- 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](https://doi.org/10.1161/CIRCGEN.117.001952).
- 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](https://doi.org/10.1111/eci.12023).

Abbreviations and units

Abbreviations

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
BSA	bovine serum albumin
CAD	coronary artery disease
cAMP	cyclic adenosine monophosphate
cCRE	candidate cis-regulatory element
CNN1	calponin 1
CO₂	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
CVD	cardiovascular disease
DAPI	4',6-diamidino-2-phenylindole
cDNA	complementary DNA
DNA	deoxyribonucleic acid
dNTP	deoxyribose nucleoside triphosphate
Dock2	dedicator of cytokinesis 2
DTT	reverse transcription
EC₅₀	half maximal effective concentration
ECAR	extracellular acidification rate
ECM	extracellular matrix
ef	enrichment factor
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

(Continued)

FCCP	carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FEV₁	forced expiratory volume in 1 second
FTP	file transfer protocol
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GES analysis	gene set enrichment analysis
GI	gastrointestinal
GWAS	genome wide association study
H3K27ac	histone 3 lysine 27 acetylation
H3K4me3	histone 3 lysine 4 trimethylation
H₂O	dihydrogen monoxide (water)
H₂O₂	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
IF	immunofluorescence
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 metalloproteinase
MMP9	MMP 9
NAC	N-acetylcysteine
O₂	elemental oxygen
O₂^{•-}	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

Abbreviations and units

(Continued)

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
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
WHO	world health organization

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

Supplement

S.1 CellROX™ assay

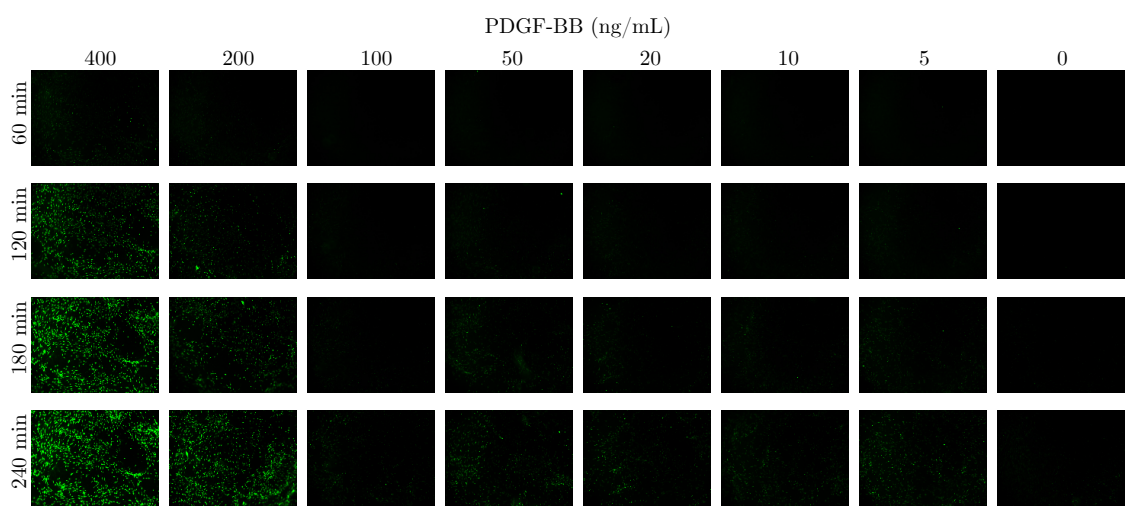


Figure S.1: Representative CellROX™ Green signal for PDGF-BB titration

S.2 Evaluation of oxidative stress with an anti-8-oxoguanine antibody

Antibodies

Name	Species	Manufacturer
Anti-8-Oxoguanine Antibody, clone 483.15	mouse	Sigma-Aldrich Co. LLC.
Alexa Flour™ 568 anti-mouse IgG (H+L)	donkey	Invitrogen™ AG

Chemicals

Name	Manufacturer
Acetone	Carl Roth GmbH + Co. KG
DAPI	Sigma-Aldrich Co. LLC.
Methanol ($\geq 99\%$)	Carl Roth GmbH + Co. KG
Triton [®] X-100	Thermo Fisher Scientific Inc.

In immunofluorescence (IF) fluorescently labeled antibodies are used to specifically target and label a molecule in a biological sample. An anti-8-oxoguanine antibody was used as a complementary approach to CellROX[™] assay. 8-oxoguanine is a common base lesion that is caused by the oxidation of guanine in the presence of ROS (Leon et al., 2016).

For the assay, HAoSMCs differentiated according to section 3.1 on Col I matrix were washed with PBS and boosted for 90 min with 500 mL M231 supplemented with 1% FBS and 200 ng/mL PDGF-BB. After, the cells were fixated with 200 μ L methanol (1:1, -20 °C) for 40 min. The methanol:acetone was removed and cells were dried for 20 min. For permeabilization, the cells were incubated with 250 μ L 0.2% Triton[®]X-100, 1% BSA (in PBS) for 30 min. Further, unspecific binding sites were blocked with 250 μ L 3.5% BSA (in PBS) for 30 min. Afterward, HAoSMCs were incubated with 500 μ L of the primary antibody (1:500 in PBS) at 4 °C over night. In the morning the cells incubated with 500 μ L AF555 anti mouse (1:1000 in PBS) and nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI) (1:5000) for 1 h.

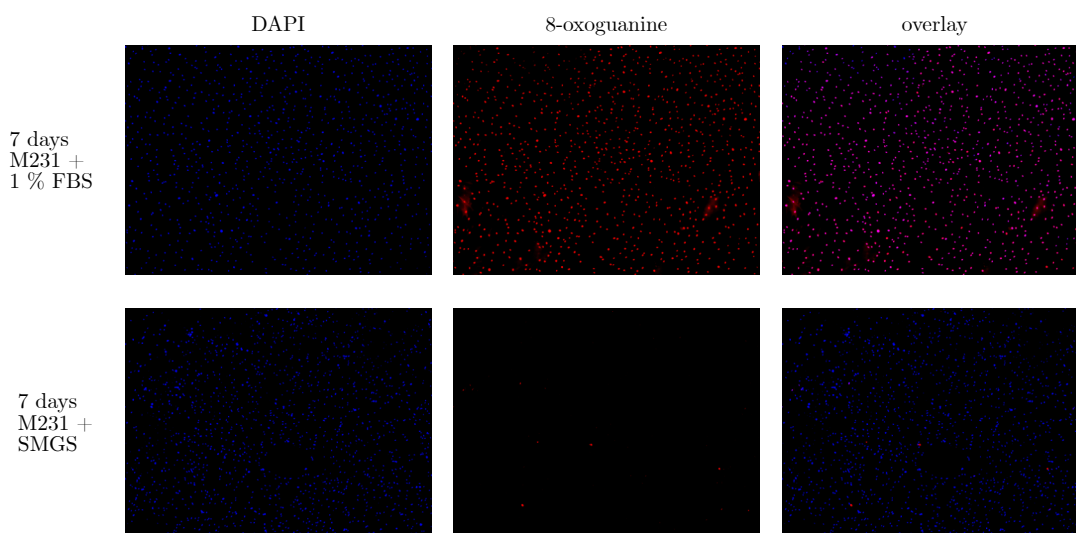


Figure S.2: Evaluation of oxidative stress with an anti-8-oxoguanine antibody

Unfortunately, this approach was unsuccessful, because, as seen in figure S.2, starvation of HAoSMCs in M231 supplemented with 1% FBS was oxidation of guanine.