INSTITUT FÜR KARDIOGENETIK

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

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

Masterarbeit

verfasst am
Institut für Kardiogenetik

im Rahmen des Studiengangs
Molecular Life Science
der Universität zu Lübeck

vorgelegt von
Torben Falk

ausgegeben und betreut von
Prof. Dr. Jeanette Erdmann

mit Unterstützung von
Dr. Tobias Reinberger

Lübeck, den 21. Juli 2022

Eidesstattliche Erklärung

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

_____

Torben Falk

Zusammenfassung

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

Abstract

Placeholder

Acknowledgements

Daaanke an alle!

Contents

1

Introduction

1.1 Coronary artery disease

CAD is among the leading causes of death in the (western) worldwide, being prevalent in about 6.7 % of American adults and killing more than 350'000 people in the USA in 2019 alone (Disease Control and Prevention, 2022; Fryar, 2012). CAD is characterized by the build-up of fatty plaques in the arteries leading to the heart. 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 can cause serious damage to the heart muscle. Next to common and well-known lifestyle factors like tobacco use or physical inactivity, CAD risk additionally has a hereditary component (Montalescot et al., 2013).

1.2 Muscle Cells in CAD

The lumen of a typical blood vessel is surrounded by three distinct layers. The outer adventitia is rich in connective tissue, shapes the vessel, and wraps the media. The media contains vascular smooth muscle cells (VSMCs) which are required to mediate vasodilation and vasoconstriction according to signaled requirements. The inner layer consists of endothelial cells that define the lumen of the vessel. (Tucker et al., 2022; Yap et al., 2021) For the longest time, the role of VSMCs in the development and progression of atherosclerosis has been underestimated and over-simplified. VSMCs have simply been considered to be either promoting arteriosclerosis or beneficial for plaque stability. Only with the emergence of novel and exciting technologies like single-cell (sc) transcriptomics and lineage tracking, this model is changing into a more multifaceted one. (Grootaert and Bennett, 2021; Yap et al., 2021) The study of VSMCs in arteriosclerosis is rapidly evolving, and the underlying models being adjusted accordingly. The black and white idea of VSMCs in arteriosclerosis existing either as a differentiated (contractile) phenotype or as a dedifferentiated (synthetic) phenotype, is making place for a model that considers a diverse set of dedifferentiated phenotypes (Grootaert and Bennett, 2021; Yap et al., 2021). The phenotypic switch describes the down-regulation 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, as well as their impact on

disease progression, are still the subject of intensive research.
Two external stimuli that seem to play central roles as cytokines determining the fate of VSMCs in atherogenesis are TGFβ & PDGF-BB.

1.3 TGFβ Signaling

TGFβ Signaling in General

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

TGFβ Signaling in VSMCs & atherosclerosis

In the context of VSMCs, TGFβ promotes proliferation and hypertrophy. Further, it promotes VSMC differentiation, via elevation of contractile gene expression as well as the down regulation 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) that is also required for phenotype switching. This way hindering (Davis-Dusenbery et al., 2011) or potentially reversing phenotype switching (Pan et al., 2020).

1.4 PDGF Signaling

PDGF Signaling in General

Five different PDGF isoforms (PDGF-AA, PDGF-AB, PDGF-BB, PDGF-CC & PDGF-DD) have been identified that are formed as dimeric combination of four distinct polypeptide chains. All five isoforms bind to tyrosine kinase receptors (platelet-derived growth factor receptor (PDGFR)α & PDGFRβ). Upon activation, the receptor 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. All these pathways are ultimately involved in the promotion of cellular proliferation, survival and migration (Chen et al., 2013; Heldin, 2013; Hu and Huang, 2015).
The predominantly expressed isoforms of endothelial cells seems to be PDGF-BB (Andrae et al., 2008; Heldin, 2013) which acts as a paracrine activator for VSMCs and other mesenchymal cells (Heldin, 2013). Signaling via PDGF-BB and the PDGFRβ plays an important role in development of multiple tissues, amongst other in the development of the cardio vascular system (Levéen et al., 1994). In the adult, PDGF-BB picks up an important role in wound healing processes (Robson et al., 1992). The role of PDGFRβ signaling in

pathologic processes like cancer or cardio vascular disease has been a subject of extensive study for decades (Heldin, 2013; Raines, 2004).

PDGF Signaling in VSMCs & atherosclerosis

In the context of VSMCs, PDGF-BB was shown to increase KLF4 levels, which results in up-regulation of mesenchymal markers as well as the loss of contractile markers. Ultimately, serving as an external stimulus for proliferation and phenotype switching (Yap et al., 2021). Similarly to the overall role of VSMCs in arteriosclerosis, the role of PDGF-BB is still the subject of extensive study. All PDGF isoforms are abundantly found in arteriosclerotic cell walls, further PDGFR expression is elevated in affected vessels (Hu and Huang, 2015). For a long time PDGF signaling and inflammation has been assumed to be disease promoting (Andrae et al., 2008; Chen et al., 2013; Hu and Huang, 2015) and in 2015 He et al. (2015) showed that PDGFRβ signaling in mouse model leads to inflammation and increased plaque formation. In contrast to this consensus, Newman et al. (2021a) were recently able to demonstrate, that sustained signaling via PDGFRβ is required for VSMC involvement in arteriosclerotic lesions. They Further observed (again in mouse model) that lack of VSMC involvement during plaque formation, can be temporarily compensated by non-VSMC-derived cells, but long-term leads to instability of arteriosclerotic lesions.

ROS in PDGF Signaling

ROS is a broad term for a class of highly reactive molecules derived from elemental oxygen (O_2). They are traditionally infamous for the damage they can do to proteins and nucleic acids when not kept in check, potentially causing irreparable damage and ultimately leading to cell death. Recently, this perception has been shifting, and especially hydrogen peroxide (H_2O_2) and superoxide anion radical ($O_2^{\bullet -}$), are being recognized for their role in cellular signaling. (Sies and Jones, 2020)
Human cells contain dozens of enzymes, which are capable of generating ROS and enzymatically maintain a steady redox state (Sies and Jones, 2020). H_2O_2 and $O_2^{\bullet -}$ serve as important second messengers in the central nervous system (Nayernia et al., 2014) or in the repair of vascular lesions (Andrae et al., 2008). Interestingly, the generation of ROS as a second messenger gets triggered by stimulation with PDGF-BB (Sundaresan et al., 1995; Bouzigues et al., 2014).

1.5 GWAS

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

GWAS

An amazing resource for getting a first glance into these interactions are GWAS, a method that allows for the identification of genetic variants associated with a phenotype.
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. After profiling the cohort on a genomic level (today

ually via microarrays, in the future most likely via whole genome sequencing (WGS)) and enotypically, the collected data needs to pass through several steps of quality control, e.g. for the removal of rare variants, miss-matched phenotypes, etc. Afterward, variants that were not directly analyzed are inferred from a reference. The final step of the initial analysis is the statistical mode regression model that is used to test for the association of all variants with the phenotype in question. 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 disease and therefore a first glimpse into the molecular biology of the observed phenotype (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, and implementing tools for clinical risk prediction. A plethora of follow-up analyses (postGWAS) can and need to be performed to determine a set of credible variants and to assess their molecular mechanism.

The first important follow-up is fine-mapping. Due to the complex linage disequilibrium (LD) of variants in the human genome (see section 1.6), loci identified in GWAS usually do not contain a single variant but form a potentially large set of linked and significant variants. Fine-mapping describes the process of identifying the actual causal variant in this ss. Multiple sophisticated statistical methods have been developed, the most popular of which is Bayesian modeling hich outputs variant-specific posterior inclusion probabilitys (PIPs) that form a credible set of potentially causal variants. It is important to remember that methods for fine-mapping are still continuously refined and will most likely keep getting better and more complex with the increasing complexity of the studied phenotypes. Further, fine-mapping is a statistical approach that will never be able to determine causality! (Schaid et al., 2018; Uffelmann et al., 2021)

After the identification of likely causal variants, the next steps aim to gain information on their effect in determinin e analyzed phenotype. Variants require mapping to impacted genes, associated pathways, and relevant tissues to get a glance at the complete picture. For these steps, no standard protocols exist and the procedure highly depends on the genomic context of the individual variant. Coding variants are rare but offer themselves to be immediately studied on a protein leve hile non-coding variants usually greatly benefit from the consultation of more high throughput data in the form of e.g. expression quantitative trait loci (eQTL) (Uffelmann et al., 2021).

Finally, the results and ideas derived from statistical models can and need to be taken k to the wet lab, to be extended and verified. Utilizing all the recent great advances in molecular and cellular biology such as the development of increasingly comprehensive *in vitro* models as well as their manipulation via methods like clustered regularly interspaced short palindromic repeats (CRISPR)-Cas gene-editing (Lichou and Trynka, 2020).

1.6 Complementary High Through Put Methods

The development of high trough put methods as well as the great increase in computing power over the last few years have spawned a plethora of incredible datasets that already have been and can be further utilized for postGWAS analyses. 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 from populations 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. p_{AB} is the frequency of the AB haplotype.
The LD becomes important 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

Problems of interpretation of GWAS data are already described in section 1.5. Link to gene (L2G) scores are an attempt at overcoming 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 output L2G scores are calibrated to represent the probability $(0, 1)$. (Mountjoy et al., 2021)

Regulatory Build

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

ENCODE cCRE

Very similarly, the ENCyclopedia Of DNA Elements project (ENCODE) project summarizes DNA accessibility and chromatin modification data into candidate cis-regulatroy elements (cCREs). Regions showing high DNase signal are further annotated to be proximal

enhancer-like elements (pELS) or distal enhancer-like elements (dELS), promoter-like elements (PLS), other regions with high histone 3 lysine 4 trimethylathion (H3K4me3) signal (which might represent poised or non-canonical promotors), or CTCF-only elements based on the existence of H3K4me3, histone 3 lysine 27 acetylation (H3K27ac), or CTCF marks. (Moore et al., 2020)

ATAC-seq

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a method to access chromatin accessibility in the genome. ATAC-seq utilizes the hyperactive Tn5 transponase to insert sequencing adapters into accessible regions of chromatine. DNA is purified and amplified via polymerase chain reaction (PCR) and then sequenced. Mapping sites with insertions in the genome allows for the identification of highly accessible genomic regions. (Buenrostro et al., 2013; Buenrostro et al., 2015a)
PCR amplification of the DNA makes this method extremely sensitive. 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 tagged with barcoded primers during the PCR. These barcodes allow mapping of ATAC-seq data to the isolated cells. (Buenrostro et al., 2015b)

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, as well as 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).

Hi-C & TADs

Hi-C is a method for mapping chromosomal conformation. To achieve this, 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, this way covalently linking DNA fragments, which were originally 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, topologically associated domains (TADs) were identified to be a basic feature of genome organization with an average size of 880 kb (Dixon et al., 2012; Wang

et al., 2018). What makes TADs of such high interest is the fact that interactions of DNA sequences are usually confined within TADs. Tissue-specific genes and their enhancers are usually found in the middle of TADs, while the edges enrich for housekeeping genes and CTCF binding sides, which might serve as insulators between different domains (Pombo and Dillon, 2015).

1.7 Aim of the thesis

The aims of this thesis are split into two quite distinct projects that both ultimately aim to contribute to a better understanding of arteriosclerosis and CAD:

- The split role of PDGF-BB during progression of arteriosclerosis (see section 1.4), indicates that PDGF-BB signaling is neither completely beneficial nor disadvantageous to diseases but there is an optimal dosage of stimulation. Combining this theory with the fact that ROS are involved in PDGF-BB signaling and also highly associated with arteriosclerosis (Burtenshaw et al., 2019), we hypothesized, that PDGF-BB signaling may cause oxidative stress, this way contributing to disease progression. The first part of this thesis will deal with the *in vitro* characterization of PDGF-BB stimulated VSMCs and the establishment of a robust assay for oxidative stress in VSMCs.
- As most data is generated by high-throughput methods, the summary statistics from Aragam et al. (2021) are a great resource. One of the goals of this thesis is to make this data and its context easily accessible to medical researchers in the form of an interactive web-based visualization tool. The GWAS Navigator will co-visualizing GWAS summary statistcs with different annotations in the form of gene products, associated phenotypes from other GWAS or putative regulatory elements.
- Finally, the data curated for the GWAS Navigator will be used in an enrichment analysis, checking different biosamples for the overlap of disease-associated variants with regulatory elements.

  Have fun with my thesis, this still is a mess...

2

Material

2.1 Manufactors

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

Continued on next page

(Continued)

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

2.2 Celllines

Name	Celltype	Manufacturer
Human Aortic Smooth Muscle Cell (HAoSMC)	prim. human cell	?!?

2.3 Primer

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

2.4 Chemicals

Name	Manufacturer
5X First Strand Buffer	Invitrogen™AG
Antimycin A	Sigma-Aldrich Co. LLC.
BSA	?!?
CellROX™Green Reagent	Thermo Fisher Scientific Inc.
Collagen Type I, rat tail	ibidi GmbH
dNTP Mix	Applied Biosystems
DTT	Invitrogen™AG
Ethanol (99.9 %)	J.T.Baker®
FCCP	Sigma-Aldrich Co. LLC.
D-(+)-Glucose, 45 % solution	Sigma-Aldrich Co. LLC.
~~GlutaMAX~~™-I	Gibco BRL
Hoechst 33342,	Invitrogen™AG
IL-1β	PeproTech©
M-MLV RT (200 U/µL)	Invitrogen™AG
NAC	Sigma-Aldrich Co. LLC.
NaHCO3	Carl Roth GmbH + Co. KG
NaOH, 1 N	Carl Roth GmbH + Co. KG
Oligomycin	Sigma-Aldrich Co. LLC.
Roth Hexanukleotid Random-Primer	Carl Roth GmbH + Co. KG
PBS	Lonza Group AG
PDGF-BB	PeproTech©
Sodium Pyruvate (100 mM)	Gibco BRL
~~RiboLock~~ RNase Inhibitor (40 U/µL)	Thermo Fisher Scientific Inc.

Continued on next page

(Continued)

Name	Manufacturer
Seahorse XF calibrant	Agilent Technologies, Inc.
PowerUp™SYBR™GREEN Master Mix	Thermo Fisher Scientific Inc.
TGFβ	PeproTech©

2.5 Media, Supplements

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

2.6 Solutions

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

2.7 Kits

Kit	Manufacturer
Total RNA Purification Kit	Jena Bioscience GmbH

2.8 Consumables

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

2.9 Devices

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

2.10 Programs & Modules

Programs

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

Continued on next page

(Continued)

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

Python Modules

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

Frameworks

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

3

Methods

3.1 Cultivation and differentiation of HAoSMCs

For the following experiments, human aortic smooth muscle cells (HAoSMCs) were used. A cell type commonly used for the study of cardiovascular function and disease ([**Reference for this claim**]). Cells were kept at 37°C and 5% CO2 whenever possible. For differentiation, cells were at first treated with TGFβ to induce a contractile phenotype and then further stimulated with IL-1β & PDGF-BB to induce a synthetic phenotype. For more information, please check the section 1.3 & 1.4 as well as the referenced literature.

Thawing & Cultivation

For longtime storage, cells were stored in liquid nitrogen. When required, new cells (6^{th} passage) were thawed at 37°C in the water bath and transferred to a 15 mL tube. After centrifugation for 2 min at 300xg the supernatant was removed and the cell pellet was taken up in 14 mL of Human Vascular Smooth Muscle Cell Basal Medium (Medium 231) (M231) + Smooth Muscle Cell Growth Supplement (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

When reaching a maximum of 80 % confluency (approx. once a week) the medium was removed completely and cells were washed once with 5 mL of phosphate buffered solution (PBS). The washed cells were incubated with 3 mL trypsin for 4 min at 37°C before 7 mL M231 were added to the detached cells. Further, the cell suspension was transferred to a 15 mL tube and pelleted for 4 min at 300xg. Finally, supernatant was removed and the pellet resuspended in M231 + SMGS, seeding 500×10^3 cells per TC Flask T75.

Preparation of Collagen I matrix

For preparation of the collagen type I (col I) matrix (1.8 mg/mL) all the components were mixed, adding the col I last. All components were stored at 4°C and all pipetting steps were carried out on ice:

Table 3.1: col I Matrix Composition

component	concentration	volume (µL)
H20	-	38.9
M231	-	53.3
SMGS	20x	5,3
NaOH	1 M	2,7
NaHCO3	7.5 %	2.1
Col I	5 mg/mL	57.6
total	-	160

160 µL of matrix mix were transferred in each used well of a Nunc™Cell-Culture Treated Multidish 24 (24 well multidish), fully coating the bottom of the wells. For polymerization, the matrix was incubated at 37°C for at least 60 min.

Differentitation of HAoSMCs

Differentiation was carried out over a total of 7 d in the 24 well multidish. 1 mL M231 was used as the medium, supplemented with 1 % fetal bovine serum (FBS) and different 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 treated surface of the 24 well multidish.
- **Day 1:** After 24 h the medium was replaced with 1 mL M231 + 1 % FBS + 5 ng/mL TGFβ (or 1 mL M231 + 1% FBS).
- **Day 5:** The medium was replaced with 1 mL M231 + 1 % FBS + 10 ng/mL IL-1β + 10 ng/mL PDGF-BB (or just 1 mL M231 + 1% FBS).
- **Day 7:** Potentially further stimulation is described in the section of the corresponding assay.

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 yielding a crossing point when signal reaches a certain threshold. A lower quantification cycle (Cq) corresponds to a higher initial DNA concentration. (Huggett and Bustin, 2011)
Quantitative PCR (qPCR) was utilized to assess the mRNA concentration of the two reporter genes calponin 1 (CNN1) and matrix metallopeptidase 9 (MMP9) in HAoSMCs ~~differentiated as described in section 3.1~~. Using the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a reference.

RNA Isolation

RNA was isolated using the Total RNA Purification Kit. The extraction was performed according to the corresponding protocol, using the extra washing step with 700 µL 80 % ethanol and eluting with 30 µL of RNase-free water.

Reverse Transcription

For reverse transcription (RT), RNA samples were diluted to yield 10 µL of ~10 ng/µL RNA. The samples were heated for 5 min at 68°C before adding 10 µL of the RT reaction mix:

Table 3.2: Master Mix for RT

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

The reverse transcription was carried out for 60 min at 37°C, before inactivating the enzyme for 5 min at 95°C. cDNA was used for qPCR or stored at -20°C.

qPCR

Table 3.3: Sample Composition for qPCR

component	~~conentration~~	volume (µL)
SYBR GREEN ~~Master~~ Mix	1:2	3.75
Primer (forward + reverse)	5 pM (each)	1.125
H20	-	1.125
cDNA	-	1.5

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

Table 3.4: qPCR Cycle

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

Processing of Data

The Cq was automatically calculated by the software SDS2.2.2 and exported for further analysis. The arithmetic mean of ~~three 3 technical~~ was calculated for each sample, disregarding values that are obvious outliers. For normalization, the mean Cq of the reference gene GAPDH was subtracted from the mean Cq of the gene of interest:

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

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

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

In total, four biological replicates were done. Data visualization and statistical analysis were done in python. Assuming a normal distribution, a student's t-test was used, and a p-value of 0.05 is considered significant. For detailed information, please refer to the script.

3.3 Energy Profiling

The Seahorse XF Analyzer allows real-time measurement of dissolved oxygen and protons in a confined small volume by using solid-state sensor probes. These are used to calculate the oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) of a cell monolayer. The OCR and ECAR are indicators for mitochondrial respiration and glycolysis respectively and can be used to assess the metabolic function of cells. (AgilentTechnologies, 2022)
Seahorse Assay was utilized to assess the energy profile of HAoSMCs differentiated as described in section 3.1. For this assay, cells were not differentiated in a 24 well multidish but an XF24 Cell Culture Microplates. Since the confined volume required for the assay would not fit the matrix, cells were cultivated without.

Seahorse Assay

On the day before the assay, the Seahorse XF Analyzer was turned set up to calibrate. The XF24 Extracellular Flux Assay Kit cartridge was left to equilibrate in Seahorse XF calibrant overnight at 37°C (in a non-CO2 environment).
On the day of the assay, cells were washed with 500 µL PBS each and afterward incubated with 500 µL XF BASE medium, supplemented with 1 mM sodium pyruvate, 10 mM glucose, 2 mM glutamine & 90 µM NaOH. The cells were left to incubate for 1 h at 37°C in a non-CO2 environment. During this time toxins for disruption of the respiratory chain were prepared and loaded into the XF24 Extracellular Flux Assay cartridge.

Table 3.5: Toxin Concentrations for XF24 Extracellular Flux Assay

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

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

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

Finally, the medium was removed and cells were stained for 15 min with 1 µg/mL Hoechst 33342 in PBS and photographed 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 by the XF Analyzer and normalized using the cell count and the signal in the control wells. In total, three biological repeats were recorded. One of which was excluded because no changes in OCR and ECAR could be detected and cells detached from the bottom of the wells during staining. For the remaining two replicates, the ~~least fitting~~ of the 5 technical repeats for each condition was manually excluded. 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. Assuming a normal distribution, a student's t-test was used, and a p-value of 0.05 is considered significant. ~~For detailed information, please refer to the script.~~

3.4 Oxidative Stress Assay

CellROX™ Green is a fluorescent dye that gets oxidized in an environment of oxidative stress ~~and then binds to DNA, showing bright-green fluorescence (Thermo~~ Fisher Scientific

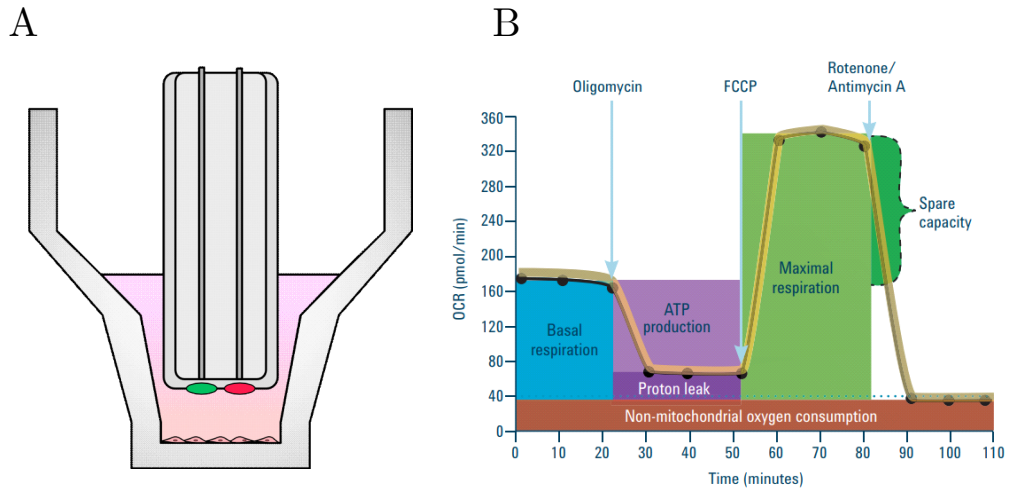

Figure 3.6: Basics of Seahorse Assay (a placeholder)
(A) Schematic of a ~~well-used~~ for Seahorse Assay. For the measurement, the piston in the middle lowers to the bottom~~, this way defining a restricted space at the bottom~~. OCR and **eacr!** in this volume are measured via two probes (red and green).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

Inc., 2022). ~~CellROX™~~ Green assay was used to assess generation of ROS in HAoSMCs ~~differentiated as described in section 3.1. After differentiation, further stimulation (from here on referred to as *boost*) with PDGF-BB was carried out.~~ Finally, a recovery experiment was performed using N-acetylcystein (NAC), a potent antioxidant, to quench generation of ROS.

CellROX™ Assay

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

Table 3.7: Composition for Seahorse Assay Boost

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

Cells were kept at 37°C in a 5 % CO_2 environment during the boost, the incubation ~~time is indicated with the results of the respective experiment~~. Imaging was done with

the ~~BZ-X810~~ All-in-One Fluorescence Microscope, using standard sensitivity. Images for the NAC quench were recorded as a z-stack and merged into one image using [KEYENCE SOFTWARE].

Processing of Data

For PDGF-BB-boost titration, ~~7~~ biological repeats were performed, ~~of which one~~ was excluded because of a high signal in the negative control. For NAC quench, ~~4~~ biological repeats were performed, ~~of which one~~ has been excluded because no signal in the positive control. For quantification of signal intensity, pixels with a green value higher than 90 were counted. Differences in cell count were adjusted by ~~division through~~ the number of pixels with a blue value bigger than 80. To adjust for the large variance in total signal intensity between biological repeats, values were adjusted by division through the total signal of all recorded conditions. For statistical testing, the Mann-Whitney U test was used, and a p-value of 0.05 is considered significant. ~~For detailed information, please refer to the scripts.~~

3.5 Curation of Data for postGWAS Analyses

Data for postGWAS analyses and co-visualization with the GWAS data, were downloaded from public resources. Processing of the data and further annotation is briefly ~~described in the following listing.~~ The generated tables are summarized in figure 4.9 and table 4.8. ~~For a complete view, please refer to the download scripts.~~

- **GWAS Summary Statistics:** The CAD GWAS summary statistics from Aragam et al. (2021) as well as a list of identified proxy single nucleotide polymorphisms (SNPs) from the study 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 43135 approved symbols, mapping to their HGNC ID as well as a list of all 98723 symbols (approved, alias, and previous), mapping to their HGNC ID.
- **Linked SNPs** LD r^2 values for variants in a 500 kb window around all variants in the list of CAD GWAS proxy variants, were computed and downloaded 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 449770 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 ~~Annotatation~~** The newest Ensembl build (Ensembl release 106) was downloaded via the ensembl FTP server. Features annotated as genes of the type protein-coding (19994), lncRNA (17734), or miRNA (1877) 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, containing 110623 open chromatin regions, 30873 TF binding sites, 175885 CTCF bindsing sites, 127935 enhancers, 36597 promotors & 140548 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 that is approved by the HGNC were dropped, yielding a total of 3580206 database entries.
- **TSS** 35160 transcription start sites (TSS) for protein-coding genes were extracted from a University of California Santa Cruz (USCS) Genome Browser dump.
- **Associated traits from GWAS catalog** The SNP trait associations from the latest release of the GWAS catalog as well as the accompanying list of studies were downloaded via the GWAS catalog FTP server. 14892 SNP-trait correlations missing a position on the human reference genome or the p-value for the association were ~~dropped~~ from the data set. Further, the column for Odds Ration or beta was separated into two columns. In total, 370002 associations from 5831 distinct studies were collected.
- **TADs** TADs predicted by software adapted from Dixon et al. (2012) were downloaded via the 3D genome browser. In total, TADs in 40 distinct biosamples were downloaded.
- **scATAC-seq from Newman et al. (2021b)** Processed scATAC-seq data for 8 cell-types [SOME MORE INFO] were scraped from the Miller Lab GitHub repository.
- **scATAC-seq from CATlas** Processed scATAC-seq data was scraped from the Ren Labs website for 222 biosamples.
- **ABC model** The ABC model data for 131 biosamples was downloaded from the Engreitz Lab FTP server. The data was further ~~translated~~ from Genome Reference Consortium Human Build 37 (hg19) to Genome Reference Consortium Human Build 38 (hg38) using pyliftover.
- **ENCODE cCREs** cCREs in distinct biosamples were downloaded by Dr. Tobias Reinberger, filtering out elements that were annotated as *unclassified*.

3.6 Visualization of GWAS data

For visualization of the data, a ~~bokeh~~ application was built, that fetches the data from the database and ~~renders~~ it to a web browser.
~~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 running in the browser. The python side of bokeh creates python objects which are serialized into JavaScript Object Notation (JSON) data and handed over to bokehJS which deserializes them into JavaScript objects that are rendered to the browser. The integrated bokeh server additionally offers the possibility to synchronize data between the underlying python environment and browser-side JavaScript library, allowing real-time updates to the displayed data.

According to good design principles, the concerns of the application are split into two

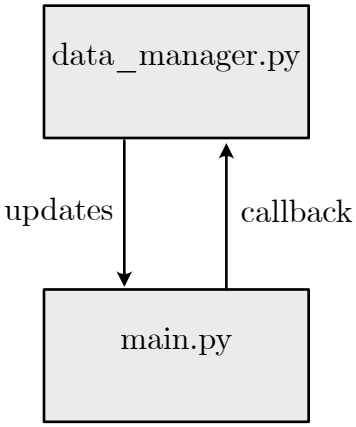

Figure 3.9: Architecture of the GWAS Navigator

sections, as shown in fig. 3.9. Reading of 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. Since data visualization, ~~as well as~~ 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 the creation of all plots and widgets as well as ~~listening for inputs~~.

3.7 Enrichment analysis

Based on the data in the database, initial postGWAS studies were run. Annotation enrichment analyses are a popular tool for the identification of terms that are over-represented in a list of interest. The most prominent application is their application as **gsea!** (**gsea!**). **gsea!**s 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, the method is used to determine ~~if~~ cCREs ~~overlaps~~ with CAD associated SNPs is enriched in certain biosamples, using Fisher's exact test.
For the analysis, cCREs annotated as unclassified were excluded. As a list of CAD associated SNPs the list of 241 proxy variants from the database was used, as well as all linked variants ($r^2 \geq 0.6$) in the 1000 ~~Genomes European Population~~. The following parameters were calculated for all biosamples:

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

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

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

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

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

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

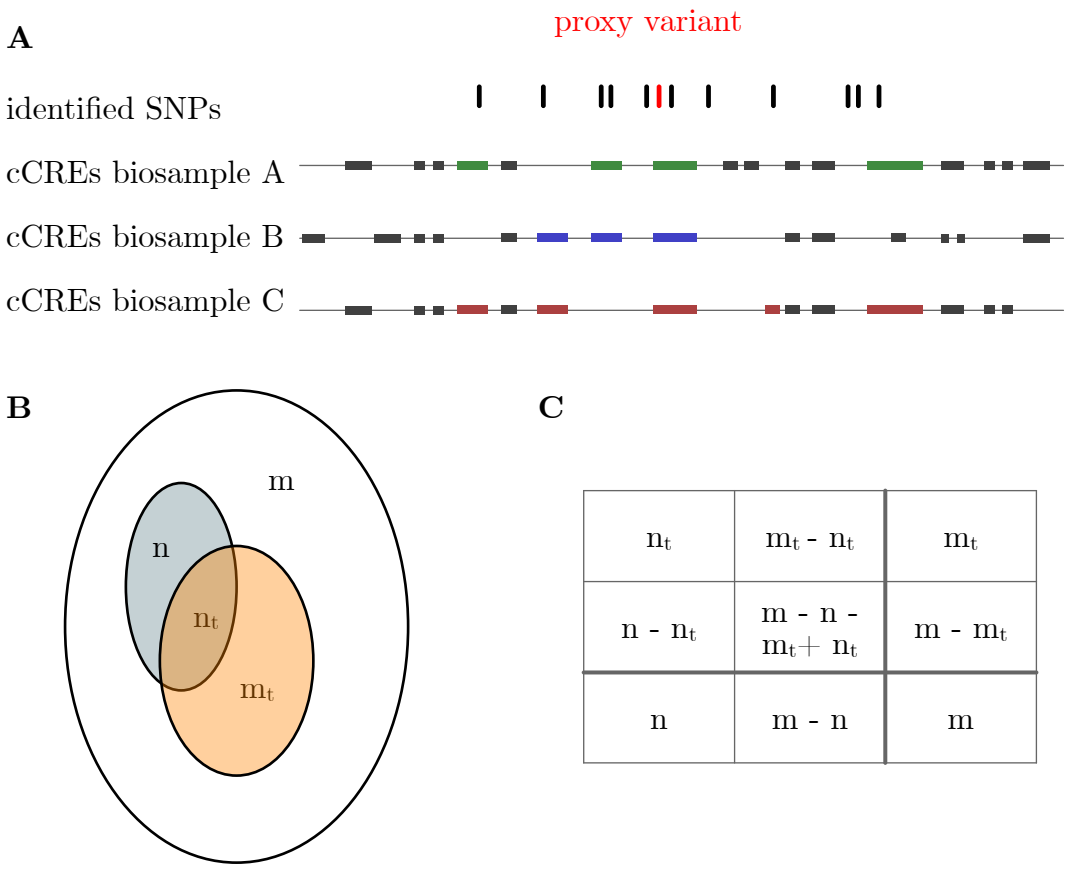

Figure 3.10: Enrichment analysis for cCREs!s overlapping with CAD risk SNPs
(A) Visual representation of the overlap calculation for enrichment calculation. The proxy variant is indicated as a red line, variants in LD are indicated as black lines. cCRE are shown as boxes, those that are overlapping with an SNP were colored according to the biosample they were annotated in. **(B)** Venn diagram of these values for a biosample. **(C)** Schematic contingency table for a biosample. (m) is the number of distinct cCREs found among all biosamples (23 in this example); (mt) the number of distinct cCREs annotated in the biosample of interest (~~16 for biosample A, 14 for biosample b, 14 for biosample C~~); (n) the number of distinct cCREs overlapping with an SNP (6 in this example); the number of distinct cCREs overlapping with an SNP in the biosample of interest (4 for biosample A (green), 3 for biosample B (blue), 5 for biosample C (red))

4

Results

4.1 Differentitaion

To characterize the influence of PDGF-BB stimulation on HAoSMCs, the cells were first treated with TGFβ for ~~2 d~~ to push them towards a phenotype that resembles the contractile phenotype. From this standardized starting point, cells were stimulated for ~~4 d~~ with IL-1β and PDGF-BB. The induced phenotypes were then characterized via qPCR and Seahorse Assay.

Expression of CNN1 & 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 ~~as well as~~ MMP9 were determined using qPCR. CNN1 as a contractile marker and MMP9 as a marker for a synthetic phenotype. For better comparability, mRNA ~~levels are considered in relation to the housekeeping gene GAPDH.~~
~~As seen in figure 4.1~~ (top panel), stimulation of HAoSMCs cultivated on a col I-matrix with TGFβ causes a significant increase in CNN1 expression ~~(+ – vs. – –)~~. After further stimulation with PDGF-BB & IL-1β, while not significant, CNN1 expression declines again (+– vs. ++) but is still significantly higher than in HAoSMCs which were not stimulated (– – vs. ++). A similar trend is noticable for HAoSMCs cultivated on plastic, ~~even when not~~ significant after four biological repeats. Additionally, stimulation of HAoSMCs on plastic with TGFβ followed by stimulation with PDGF-BB & IL-1β, ~~yields a significantly higher expression of CNN1 (++ Matrix vs. ++ Plastic).~~
As seen in the bottom panel of figure 4.1, after 4 biological repeats, no statistically significant trends can be observed for the expression of MMP9. Still the average expression of MMP9 seems to be ~~doubled for all conditions~~, the most prominent difference being between HAoSMCs treated first with TGFβ, then with PDGF-BB & IL-1β (++ Matrix vs. ++ Plastic, $p = 0.063$).

Energy profile

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

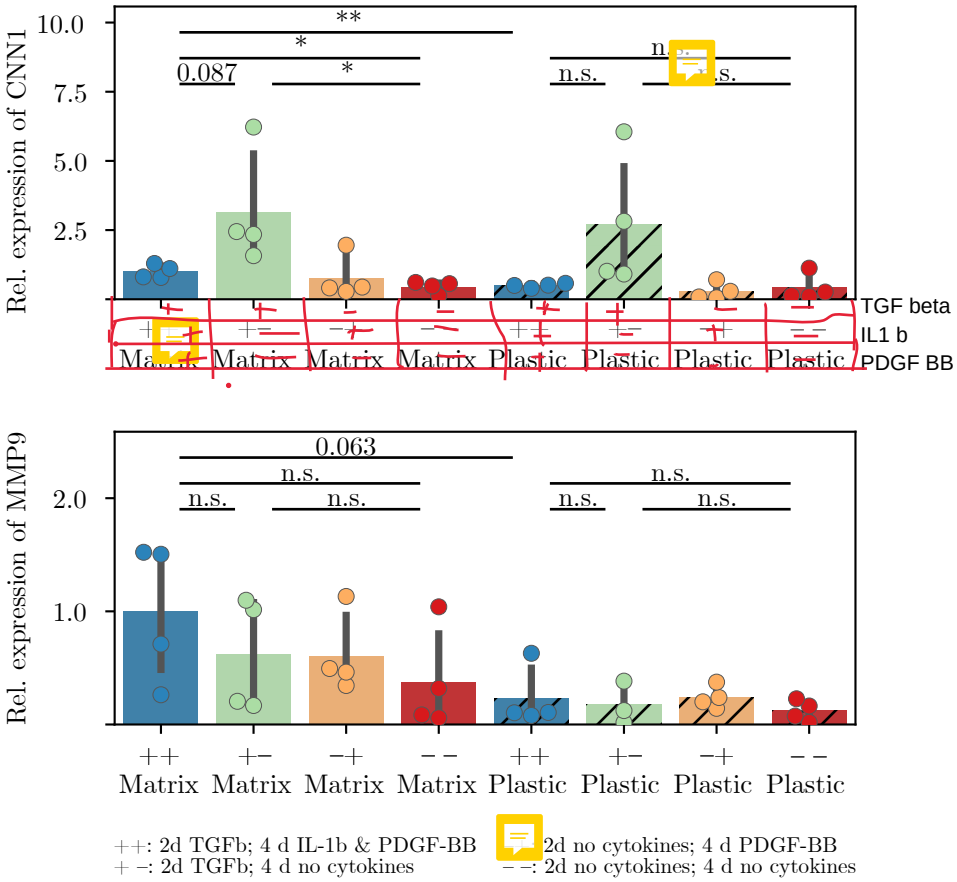

Figure 4.1: Relative Expression of CNN1 & MMP9 in HAoSMCs
qPCR analysis of expression for contractile marker CNN1 (top) and synthetic marker MMP9 (bottom) for HAoSMCs differentiated with different combinations of cytokines: **++:** 2 d with TGFβ followed by 4 d with IL-1β & PDGF-BB; **+−:** 2 d with TGFβ followed by 4 d without stimulation; **−+:** 2 d without stimulation followed by 4 d with IL-1β & PDGF-BB; **—:** 6 d 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: $* : p < 0.05; ** : p < 0.01$

plastic because the col I matrix does not fit into the confined compartment created by the piston for detection of OCR & ECAR. ~~Further~~, only two biological repeats were evaluated, because it became increasingly clear, that all other experiments would be carried out on a col I. Therefore all the following results should ~~be considered under these circumstances.~~

The readout parameters of the Seahorse assay are the OCR as a representation of mitochondrial activity and the ECAR, representing the glycolytic activity of the cells. OCR and ECAR for HAoSMCs are displayed in figure 4.2. All cells show characteristic changes in OCR after addition of toxins impacting the respiratory chain (compare to figure 3.6 B). After inhibition of the ATP synthase with Oligomycin, the basal OCR drops, this way ~~making~~

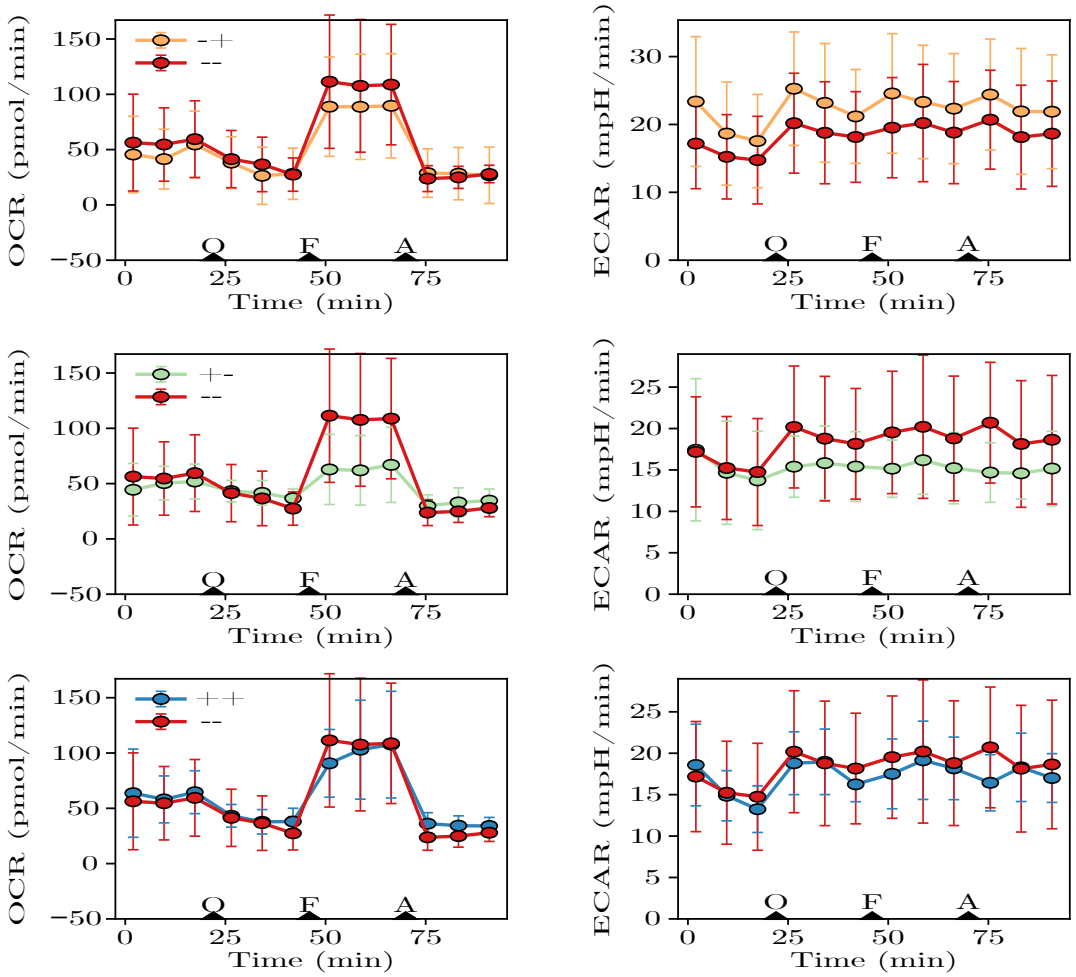

Figure 4.2: OCR & ECAR of HAoSMCs
Seahorse assay for HAoSMCs differentiated with different combinations of cytokines. **++:** 2 d with TGFβ followed by 4 d with IL-1β & PDGF-BB; **+−:** 2 d with TGFβ followed by 4 d without stimulation; **−+:** 2 d without stimulation followed by 4 d with IL-1β & PDGF-BB; **——:** 6 d without stimulation. OCR & ECAR are shown for −+ (top), +− (middle) and ++ (bottom) in comparison to ——. Injectiontimes 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 proportion of the OCR ~~accessible~~ that was required for adenosine triphosphate (ATP) production. ~~Further~~, the addition of carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP) decouples the respiratory chain~~, destroying~~ the proton gradient over the mitochondrial membrane ~~and letting the~~ cells reach their maximal respiratory capacity. Finally, the inhibition of coenzyme Q-cytochrome c reductase (complex III) with Antimycin A, stops all mitochondrial respiratory activity.
The ECAR shows a mild increase for all conditions after the addition of Oligionmycin, most likely because the cells are compensating for the loss of mitochondrial ATP production via

increased glycolysis.

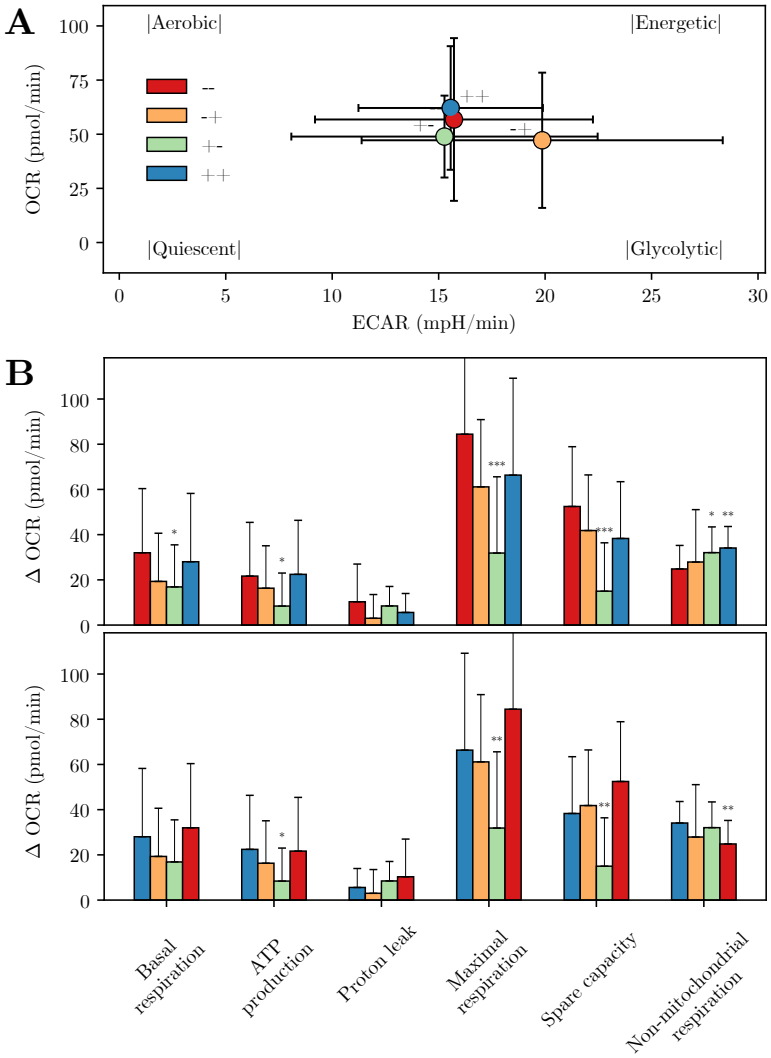

Figure 4.3: Energy profile of HAoSMCs
Seahorse assay for HAoSMCs differentiated with different combinations of cytokines as described in figure 4.2. (**A**) Initial OCR & ECAR of the four tested conditions. (**B**) Characteristics of the respiratory chain calculated from the tracks shown in figure 4.2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

Looking at the energy profile of the HAoSMCs it is easy to see that OCR & ECAR are quite similar for the conditions ++, +– and —. The only outlier showing a higher ECAR, are HAoSMCs only stimulated with only IL-1β & PDGF-BB (–+) (fig. 4.3, A). More interesting differences can be observed when examining characteristics of the respiratory chain. Stimulation with only TGFβ causes a significant decrease in basal respiration, ATP production, maximal respiration as well as spare capacity (figure 4.3, B top). Further stimulation with IL-1β & PDGF-BB then causes a significant increase of these parameters

to similar levels as in initially dedifferentiated HAoSMCs (figure 4.3, B bottom).

4.2 Evaluation of oxidative Stress

Finally, it was evaluated if further stimulation with PDGF-BB would ~~yield~~ the generation of ROS to an extent that can not be compensated by the ROS defense and lead to oxidative stress.

PDGF boost of out cells induces oxidative stress

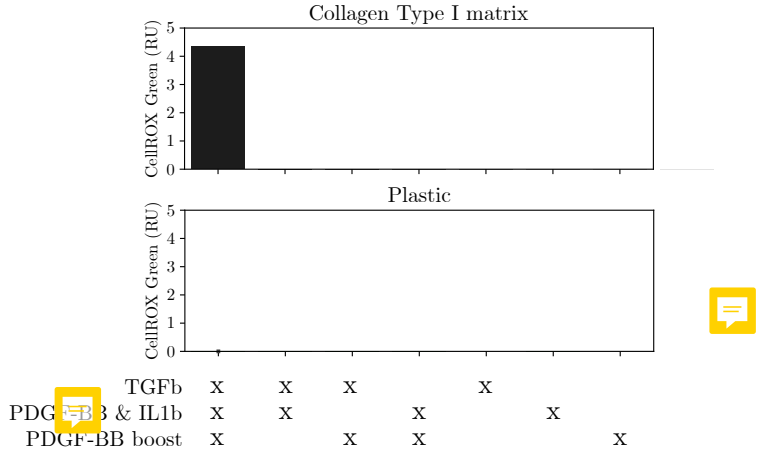

Figure 4.4: Boost with PDGF-BB induces generation of ROS.
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: 2 d with TGFβ; followed by 4 d with IL-1β & PDGF-BB; followed by 2 h boost with 200 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix (top) or plastic (bottom). The hown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by DAPI signal. No statistical analysis for (n = 1) biological repeats was performed.

~~At first, an experiment already done in the group was repeated.~~ 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 2 d with TGFβ, followed by 4 d with IL-1β & PDGF-BB, followed by a 2 h boost with PDGF-BB, was able to trigger noticeable generation of ROS for cells cultivated on col I-matrix. No generation of ROS was detectable for HAoSMCs cultivated without the col I-matrix.

Characterization of the CellROX™ Assay

To get a better understanding of the assay and its limits, a titration was carried out. For this, HAoSMCs stimulated for 2 ~~d~~ with 5 ng/mL TGFβ as well as 4 d with 10 ng/mL IL-1β & 10 ng/mL PDGF-BB, were boosted with different concentrations of PDGF-BB (0 - 400 ng/mL). ~~Signal was~~ detected after 60, 120, 180 ~~&~~ 240 min in HBSS. ~~As seen in figure 4.5,~~ CellROX™ Green signal is negligible after 60 min and then increases with elongated

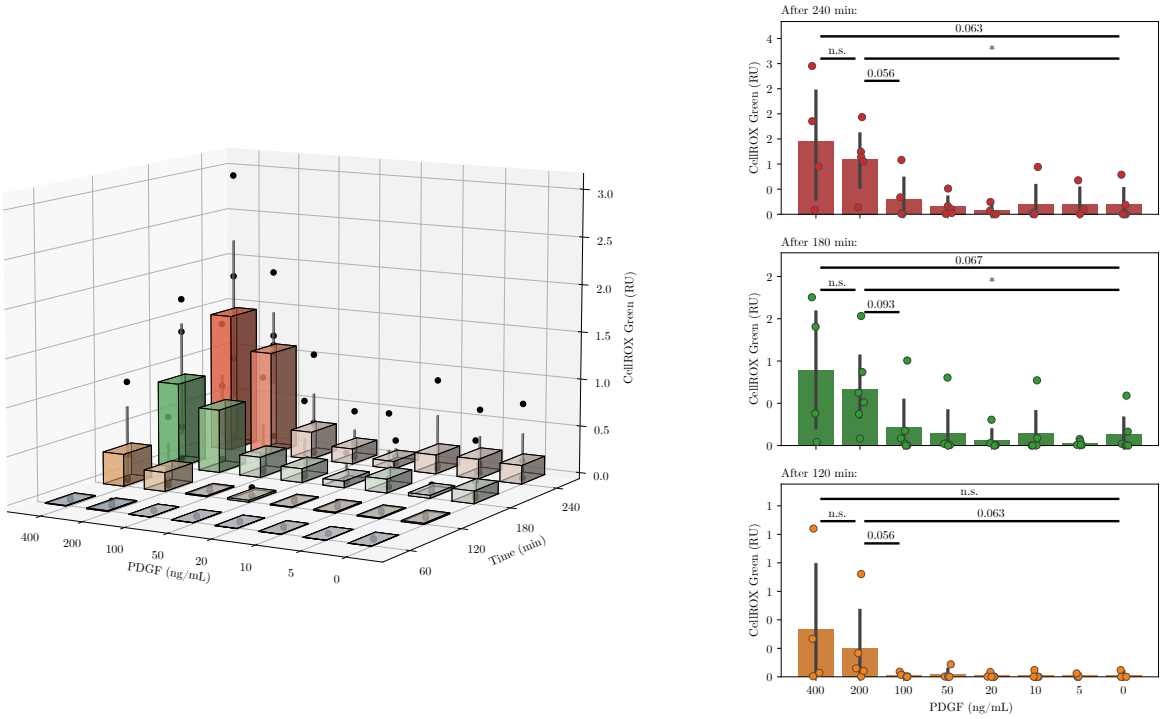

Figure 4.5: PDGF-BB boost titration
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: 2 d with TGFβ; followed by 4 d with IL-1β & PDGF-BB; followed by 4 h boost with 0 - 400 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix. (**A**) 3D visualization: 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 & 240 min. The shown signal was calculated according to section 3.4 as the CellROX™ Green signal, normalized by DAPI signal. Statistical analysis for (n = 6) biological repeats was performed using Mann-Whitney U test: $* : p < 0.05; ** : p < 0.01$. ~~It is important to note that not for every biological repeat,~~ *all* PDGF-BB concentration were tested.

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

Overall, the trend of greatly increased CellROX™ signal for boost with 100 as well as 200 ng/mL PDGF-BB was consistent within biological repeats, however variance between repeats was almost as high as differences between the conditions. Potential causes for this phenomenon are discussed in section **??**. To account for this large variation between biological repeats, the assay was reevaluated by the selection of shared conditions among the biological repeats, that were normalized ~~by~~ the cumulative intensity of all conditions

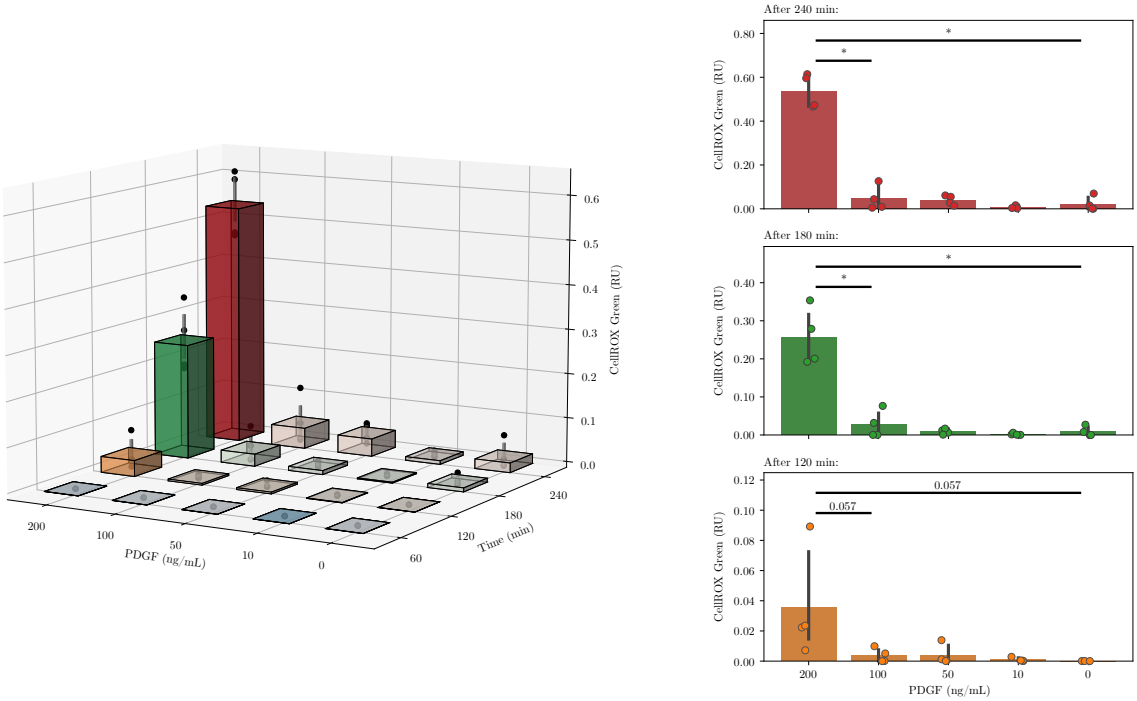

Figure 4.6: PDGF-BB boost titration - normalized
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: 2 d with TGFβ; followed by 4 d with IL-1β & PDGF-BB; follwoed by 4 h boost with 0 - 200 ng/mL PDGF-BB. Differentiation and assay carried out on col I matrix. (**A**) 3D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

of the biological repeat (see figure 4.6). This way compensating for differences between biological repeats. ~~The observations stays the same:~~ CellROX™ Green signal after 180 and 240 min is significantly higher for cells boosted with 200 ng/mL PDGF-BB than cells that were not boosted (0 ng/mL PDGF-BB).

Rescue of ROS production using NAC

Finally, a rescue experiment was performed, to verify that the observed signal in the CellROX™ assay was indeed due to the generation of ROS. For this, ROS generation was quenched by the addition of 2, 4, or 8 mM of NAC. Indeed, ~~while not statistically significant after 3 repeats,~~ a clear trend can be observed: HAoSMCs treated with NAC show no signal.
~~In the end, it should be noted,~~ that the signal only builds up over 15 - 20 min under the

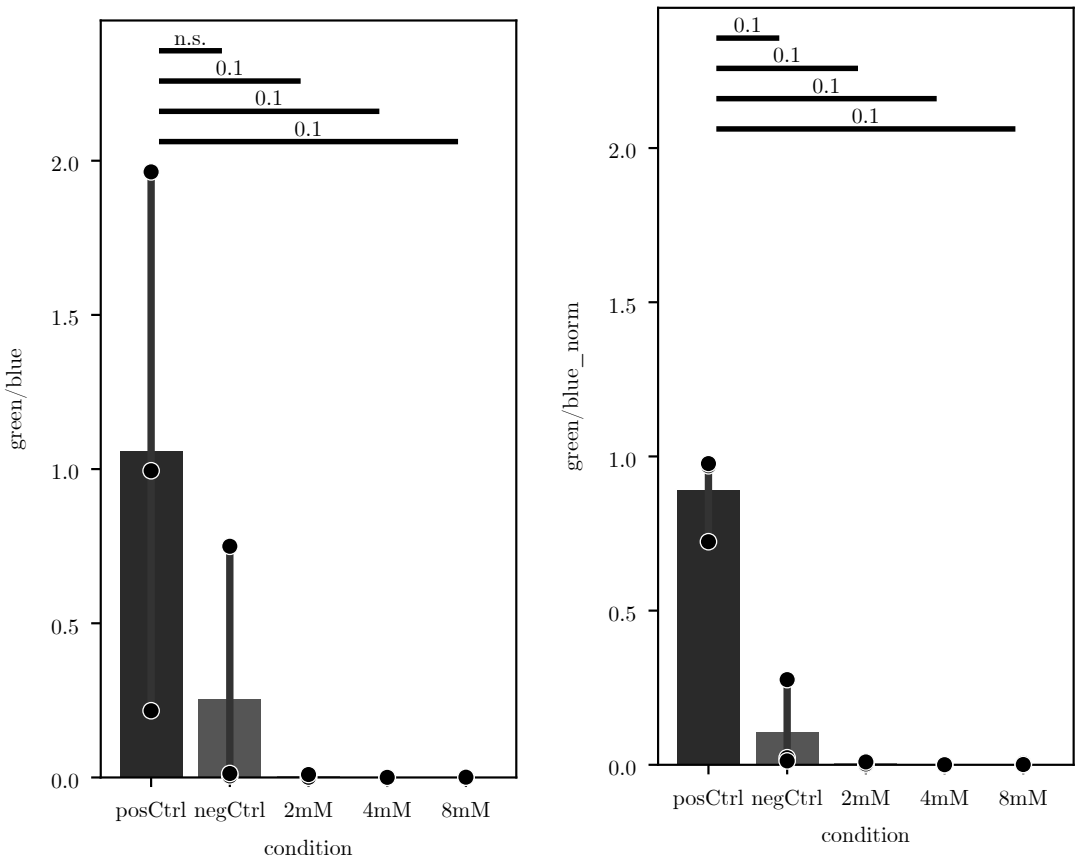

Figure 4.7: ROS generation due to PDGF-BB boost can be rescued with NAC
CellROX™ assay for HAoSMCs differentiated with different combinations of cytokines: 2 d with TGFβ; followed by 4 d with IL-1β & PDGF-BB; ~~follwoed~~ by 3 h boost with 200 ng/mL PDGF-BB. Differentiation and assay carried out on col 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 DAPI signal (**A**), further the signal was normalized via the total signal of the biological repeat (**B**). Statistical analysis for (n = 4) biological repeats was performed using Mann-Whitney U test: $* : p < 0.05; ** : p < 0.01$. pos Ctrl: not treated with NAC, negCtrl: no boost with PDGF-BB

microscope after the cells were taken out of the incubator. This indicates that generation of ROS might not be exclusively triggered by PDGF-BB boost but could also require additional contributors like the loss of the optimized atmosphere of 37°C and 5 % ~~CO2~~ in the incubator. This might not have ~~been noted~~ during the titration assay, because cells were taken out of the incubator after one hour to image them for the first time.

4.3 Database and GWAS Visualizer

Curation of Data

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

Data	Tables	Page count (including indices)
GWAS Summary stats	variation	418318
	gwas_meta_cad	867025
	identified_proxy_SNPs_tbl	4
HGNC gene list	hgnc_all_symbols_tbl	826
	hgnc_approved_symbols_tbl	592
Linked SNPs	linked_SNPs_tbl	8819
	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	8778
	ensembl_reg_build_features_tbl	1
TSS	tss_tbl	481
Open Target Genetics Scores	opentarget_l2g_tbl	40984
GWAS catalog	gwascatalog_associations_tbl	10569
	gwascatalog_studies_tbl	326
TADs	tad_tbl	902
	tad_sample_tbl	1
scATAC seq textcite{}	clint_miller_tbl	12370
	clint_miller_biotypes_tbl	1
scATAC seq CATlas	catlas_tbl	308574
	catlas_biotypes_tbl	3
ABC model	abc_tbl	153920
	abc_targetgenes_tbl	84
	abc_celltypes_tbl	3
	abc_classes_tbl	1
ENCODE cCREs	ENCODE_CCRE	4451476
	ENCODE_CCRE_META	107
total	-	6284781 (≈ 25.75 GB)

The first step towards visualization of GWAS data and postGWAS studies, was the curation of suitable complementary data. Datasets from a diverse set of data sources were downloaded and funneled into an SQLite3 database ~~according to~~ section 3.5. ~~A SQL database is a two-dimensional relational database that allows easy and fast access to the data for visualization purposes.~~ The types of data and their applications are briefly described in section 1.6. All database tables as well as their sizes are summarized in table 4.8.

The relationships between the tables as well as fields serving as a primary key, foreign or fields on which an index exists, are summarised in the databases entity-relationship (ER) diagram in figure 4.9.

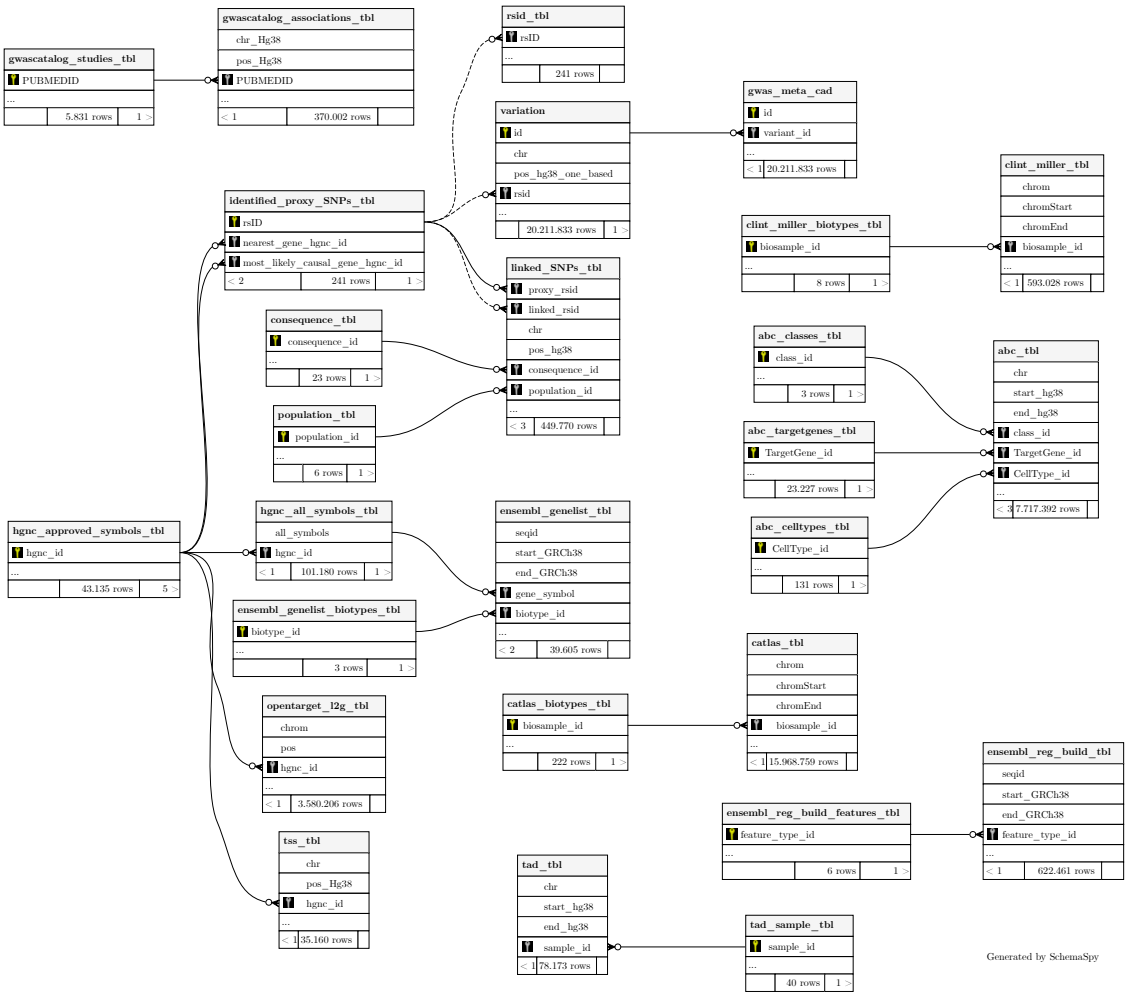

Figure 4.9: Entity-Relationship Diagram of the Database
Fields and relationships of the tables listed in table 4.8. Primary keys are marked with a ~~golden key icon,~~ fields serving as an forgein key are marked with an gray key icon. ~~On fields whichs names are spelled out but which do not have an key icon and index exists,~~ The diagram was generated via SchemaSpy.

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 is aligned with genomic elements such as genes, regulatory elements, acATAC-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 ~~split~~ search bar which has a field to specifically search for variants by their rsID as well as a field that allows searching for genes by their symbol. In case the searched gene is associated with one of the proxy variants in Aragam et al. (2021), the tool returns a list of these variants, else the tool returns the most significant variant in the proximity of the searched gene. After a variant was chosen, the tool displays the 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 Manhatten plot, showing the position of a variant on hg38 on the x-axis and its p-value on the y-axis. r^2 values of variants in LD are color-coded, ~~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). ~~Further~~, the region is aligned with protein-coding genes, lncRNAs ~~&~~ miRNAs ~~as~~ annotated in Ensembl, ~~and the~~ names of genes that are associated with the variant ~~central to analysis~~ (open target genetics L2G score > [FIND THE THRESHHOLD]) are labeled in red. In addition, regulatory elements from the Ensembl regulatory build are displayed. Finally, scATAC-seq data and enhancers-promotor links from the ABC model were aligned, automatically hiding tracks that have no elements in the visualized region. The GWAS visualizer additionally has a settings tab, in which individual tracks can be hidden.

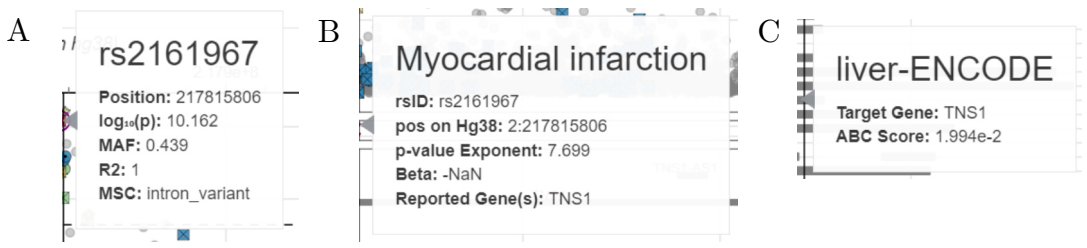

Figure 4.11: The GWAS Visualizer - Hovereffect
Exampalry hover effects for features displayed in the GWAS Navigator. (**A**) Hover for variants in the manhattan plot. (**B**) Hover for variant phenotype associations. (**C**) Hover for cell type specific enhancers in the ABC model.

4.4 Enrichment analysis

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

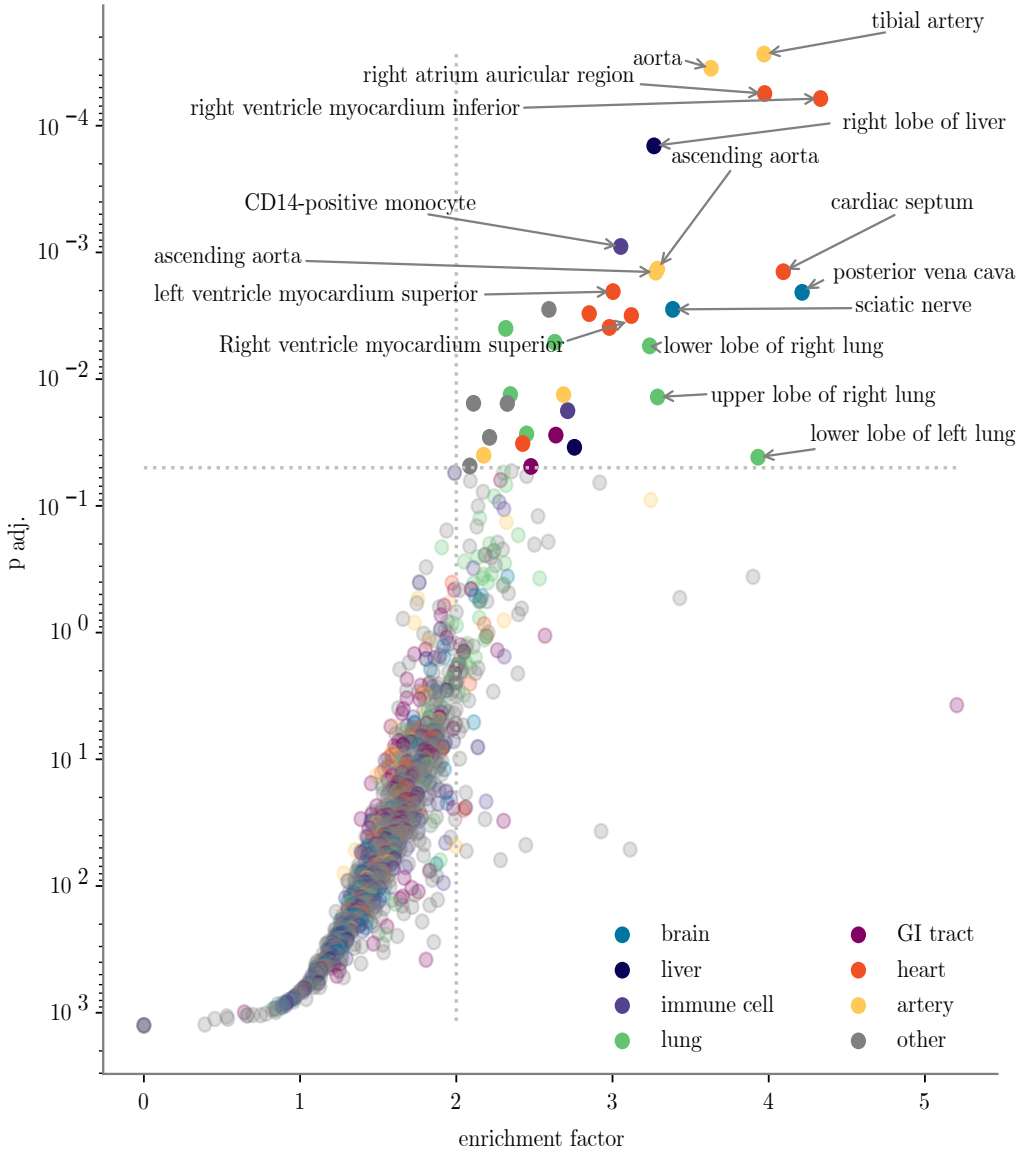

Figure 4.12: Enrichment Analysis for overlap of CAD GWAS proxy variants with tissue specific cCREs
p-values and enrichment factors for the overlap of CAD GWAS proxy variants (and variants in LD) and tissue specific cCREs. For details please refer to section 3.7.

~~As seen in figure 4.12,~~ statistical significant enrichment ($p_{adj.} < 0.05$) was observed for 34 biosamples. Using the biosample annotations from Cellosaurus, these biosamples were assigned to their tissue of origin. ~~As summarized in table 4.13 most~~ prominent groups of origin tissues were the heart (8), the lungs (7), and arteries (6). ~~Followed by~~ the liver, the gastrointestinal (GI) tract, the brain, and immune cells (CD+ monocytes).

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

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

5

Discussion

WHAT TO DISCUSS

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

- Differentiation -> conext for MMP9, CNN1. is it within the expectations, where do the expectatiosn come from, what kind of cell may we have?

5.2 CellROX Green is Suitable for Assessing ROS Generation in HAoSMCs

- CellROX -> Seems like we can prime the Cell for ROS generation. First line of evidence towards our hypothesis. - consider that taking out of the incubator might be nessecary, so maybe just priming the cells? - we require additional evidence - are these processes of biological relevance. do the concentrations make sense?

5.3 The GWAS Navigator

- We did collect a bunch of data - examplary work flow for checking STAT1 maybe

5.4 CAD Associated Variants Enrich in Regulatory Elements of Heart, Artery & Lung Tissue

- The enrichment -> do the identified tissues make sense? - immune cells - lung tissue

6

Conclusion & Outlook

We are closer to doing postGWAS analyses, we really hope that the database makes everything smoother. And we have a system where we can functionally access these identified SNPs. We are close to a point were we can combine both part of the project.

integration of ensembl database

Abkürzungsverzeichnis

24 well multidish Nunc™Cell-Culture Treated Multidish 24
ABC activity by contact
API application programming interface
ATAC-seq assay for transposase-accessible chromatin using sequencing
ATP adenosine triphosphate
CAD coronary artery disease
CNN1 calponin 1
CRISPR clustered regularly interspaced short palindromic repeats
CTCF CCCTC binding factor
Cq quantification cycle
DE Deutschland
DNA deoxyribonucleic acid
DTT reverse transcription
ECAR extracellular acidification rate
ENCODE ENCyclopedia Of DNA Elements project
FBS fetal bovine serum
FCCP carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FTP file transfer protocol
GAPDH glyceraldehyde-3-phosphate dehydrogenase
GESA gene set enrichment analysis
GWAS genome wide association study
H3K27ac histone 3 lysine 27 acetylation
H3K4me3 histone 3 lysine 4 trimethylathion
HAoSMC human aortic smooth muscle cell
HBSS Hanks balanced salt solution
HGNC Human Gene Nommenclature Consortium
H_2O_2 hydrogen peroxide
IL-1β interleukin 1 beta
JSON JavaScript Object Notation
KLF4 Kruppel-like factor 4
L2G link to gene
LD linage disequilibrium
M231 Human Vascular Smooth Muscle Cell Basal Medium (Medium 231)
MAP mitogen activated protein
MI myocardial infarction
MMP9 matrix metallopeptidase 9
NAC N-acetylcystein
OCR oxygen consumption rate
O_2 elemental oxygen

$O_2^{\bullet -}$ superoxide anion radical
PBS phosphate buffered solution
PCR polymerase chain reaction
PDGF-BB platelet-derived growth factor-BB
PDGFR platelet-derived growth factor receptor
PI3K phosphatidylinositol 3'-kinase
PIP posterior inclusion probability
PLS promoter-like elements
REST representational state transfer
RNA ribonucleic acid
ROS reactive oxygen species
RT reverse transcription
SMGS Smooth Muscle Cell Growth Supplement
SNP single nucleotide polymorphism
STAT signal transducers and activators of transcription
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
VSMC vascular smooth muscle cell
WGS whole genome sequencing
cCRE candidate cis-regulatroy element
col I collagen type I
d day(s)
dELS distal enhancer-like elements
dNTP deoxyribose nucleoside triphosphate
eQTL expression quantitative trait loci
h hour(s)
hg19 Genome Reference Consortium Human Build 37
hg38 Genome Reference Consortium Human Build 38
min minute(s)
pELS proximal enhancer-like elements
qPCR quantitative PCR
MAF minor allele frequency
GI gastrointestinal
ER entity-relationship
sc single-cell

Bibliography

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

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

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

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

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

Yap, Carmen et al. (2021). Six Shades of Vascular Smooth Muscle Cells Illuminated by KLF4 (Krüppel-Like Factor 4). Arteriosclerosis, Thrombosis, and Vascular Biology, ATVBAHA121316600. DOI: 10.1161/ATVBAHA.121.316600.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.