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, this is more or less what I am doing:

I am currently writing my Master thesis at the university of Lübeck at the Institute for Cardiogenetics on the topic of "Identification of genetic risk variants for atherosclerosis using oxidative stress assays in vascular smooth muscle cells and bioinformatic approaches":

Coronary artery disease (CAD) describes the arterial build-up of fatty deposits to a point where the blood supply to the heart gets interrupted. It is one of the major causes of death worldwide. Risk factors for CAD are typical lifestyle factors like smoking or physical inactivity, but also include hereditary factors (**cdcCoronaryArteryDisease2021**; **CoronaryHeartDisease2018**). These can provide access to the molecular pathology of the disease. One amazing resource for studying these interactions are genome wide association studies (GWAS). Unfortunately, GWAS are just the first step in a longer 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 (**lichouFunctionalStudiesGWAS2020a**).

We hypothesize that oxidative stress in smooth muscle cells plays a role in stability of atherosclerotic plaques. For this reason, I am cultivating and differentiating primary human smooth muscle cells and characterizing them using oxidative stress assay, qPCR, seahorse assay & immunofluorescence (IF).

Additionally, I am working with GWAS data on CAD (**aragamDiscoverySystematicCharacterization2021a**). Curating further publicly available data that can be used for bioinformatic follow-up analyses like the enrichment for involved tissues. Further, I am using the data to build a web application that allows co-visualization and visual exploration.

Acknowledgements

Daaanke an alle!

Contents

Abkürzungsverzeichnis

A adenin

AF488 Alexa Fluor® 488

bp basepairs

C cytosin

DNA deoxyribonucleic acid

dsDNA doublestranded DNA

E. coli *Eschericha coli*

EDTA Ethylenediaminetetraacetic acid

G guanine

HCl hydrogen chloride

qPCR quantative polymerase chain reaction

T thymin

HAoSMC human arortic smooth muscle cell

IF immunoflourescence

TGFβ Transforming Growth Factor beta

SMGS Smooth Muscle Cell Growth Supplement

IL-1 Interleukin 1 beta

PDGF-BB Platelet-derived growth factor-BB

TGFβ Transforming Growth Factor beta

DE Deutschland

MMP9 Matrix metallopeptidase 9

CNN1 Calponin 1

GAPDH Glyceraldehyde-3-phosphate dehydrogenase

col I collagen type I

OCR oxygen consumption rate

ECAR extracellular acidification rate

FCCP Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone

CAD coronay artery disease

PBS phosphate buffered solution

FBS fetal bovine serum

Cq quantification cycle

RNA ribonucleic acid

RT reverse transcription

DTT reverse transcription

dNTP deoxyribose nucleoside triphosphate

ROS reactive oxygen species

NAC N-acetylcystein

cCRE candidate cis regulatory elements

LD linage disequilibrium

Contents

1

Introduction

1.1 Coronary artery disease

coronay artery disease (CAD) is among the leading causes of death in the (western) world wide, being prevelant in about 6.7 % of US-american adults and killing more than 350'000 people in the USA in 2019 (**cdcHeartDiseaseFacts2022**; **fryarPrevalenceUncontrolledRisk2012**). Its most common complication is myocardial infarction (MI), which is characterized by chest pain (angina) and can cause serious damage to the heart muscle. It is triggered by the build up of fatty plaques in the ateries leading to the heart. This process, also called artherosclerosis, can interrupt the blood supply of the heart (**HeartAttack2017**). Known risk factors for artherosclerosis and CAD are of course life style factors such as tobacco use or physical inactivity (**taskforcemembers2013ESCGuidelines2013**), but also hereditary components.

1.2 GWAS

These hereditary factors of disease can provide access to the molecular pathology of the disease.

GWAS

- One amazing resource for studying these interactions are genome wide association study (GWAS). Orginally introduced in 2007 - GWAS a tool to correlat genetic variants with phenotypes - Experimental setup : - data collection: phenotypic and genotyping - direct studies (refined specific phenotyping) or in most cases from public biobanks - genotyping usually via micro array -> WGX is expected to become the norm - QC: example steps: - removing rare variants, variants not in Hardy Weinberg eqilibirum or with miss matched phenotypes - imputation - inference of genotypes that have not be analyzed directly by a reference - Standardized pipelines - statistical test - be aware of bias due to ancestry - regression model to test for association (linear or log depending on continous or binary phenotype) - inclusion of covariants like age, sex and ancestry.

Model:

$$YW\alpha + X_S\beta_S + g + e$$

$$q(0, \sigma_A^2 \Psi)$$

$$e\tilde{N}(0, \sigma_e^2 I)$$

Pitfalls: Bias in all steps, most importantly data collection: source of data, ethnicities.

Post GWAS

Unfortunately, GWAS are just the first step in a longer 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. - most identified stuff tells a story of a lot of variants with low individual risk - next steo is interpretation - PostGWAS

1.3 Muscle Cells in CAD

- We now that smooth muscle cells play a key role - It is widely accepted that there is not only one type of smooth muscle cell - Go into the contractile phenotype and synthetic phenotype - TGFb or PDGF used to induce them - contractile phenotype is thought to be protective

1.4 PDGF Signaling and Oxidative Stress

PDGF Signaling

ROS

- ROS in general and the role of ROS in disease

ROS in PDGF Signaling

- ROS as a second messenger in PDGF signaling

1.5 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 all ready have been and can be further utilzed for postGWAS analyses. A short overview of some definitions and methods mentioned in this thesis can be found in the follwing paragraphs:

Linkage Disequilibrium

Linkage disequilibrium is a parameter from populations genetics that describes the non-random association of two or more allels. The LD is often quantified using the correlation coefficient r^2 (**slatkinLinkageDisequilibriumUnderstanding2008**).

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

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

Where p_A and p_B is the frequency of the alleles A and B respectively. p_{AB} is the frequency of the AB haplotype.

The linage disequilibrium (LD) becomes important in the context of GWAS because identified SNPs often do no occur in isolation, but can span large haplotype blocks in the genome (**slatkinLinkageDisequilibriumUnderstanding2008**).

Locus To Gene Scores

Problems of interpretation of GWAS data are already described in section 1.2. link to gene (L2G) scores are an attempt at overcoming the challenges of establishing causal realtionships 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 (**mountjoyOpenApproachSystematically2021**). The output L2G scores are calibrated to represent the probability (0, 1) that the

Regulatory Build

The ensembl regulatory build compiles a summary of regulatory regions found in the human genome. It is build on the basis publically available epigenetic marks and transcription factor binding and contains Promotors, Proximal enhancers, distal enhancers and CTCF binding sites (**zerbinoEnsemblRegulatoryBuild2015**).

ENCODE cCREs

(**SCREENSearchCandidate**; **mooreExpandedEncyclopaediasDNA2020**)

scATAC Seq

(**buenrostroTranspositionNativeChromatin2013**) (**turnerCellspecificChromatinLandscape2021a**) (**zhangSinglecellAtlasChromatin2021**)

ABC Model

(**fulcoActivitybycontactModelEnhancer2019a**; **nasserGenomewideEnhancerMaps2021a**)

TADs

(**wang3DGenomeBrowser2018**)

1.6 Aim of the thesis

- Build tool for visual exploration of the CAD GWAS data. - Establish a system to test the role of ROS in CAD.

2

Material

2.1 Manufactors

Manufacturer	Seat
Agilent Technologies, Inc.	Santa Clara, CA, USA
Assistent - look this up	?!, DE
BRAND GMBH & Co. KG	Wertheim, DE
ChemoMetec A/S	Allerod, DK
Eppendorf AG	Hamburg, DE
Heraeus Holding GmbH	Hanau, DE
Merck KGaA	Darmstadt, DE
Keyence Corporation	Osaka, JP
Kisker Biotech GmbH & Co. KG	?!, DE
Sarstedt AG & Co.	Nürnberg, DE
Sigma-Aldrich Co. LLC.	St. Louis, MO, USA
Thermo Fisher Scientific Inc.	Waltham, MA, USA
Pepro Tech -> look this up	?!
Pechiney Plastic Packaging, Inc.	Chicago, IL, USA
J.T. Baker -> look up	?!
GFL mbH	Burgwedel, DE
AB -> look this up	Warrington, UK
Gibco BRL (Brand of Thermo Fisher Scientific Inc.	Gaithersburg, MD, USA
Ibidi -> loo this up	

(Continued)

Manufacturer	Seat
Invitrogen™ (Marke von Thermo Fisher Scientific Inc.) (Brand of Thermo Fisher Scientific Inc.)	
Lonza Group AG	Basel, CHE
Biosell -> look this up	
New England Biolabs GmbH	Ipswich, MA, USA
Nikon Corporation	Minato, JP
Sartorius ?!	Göttingen, DE
SensoQuest GmbH	G¨ottingen, DE
HANNA Instruments	??
Heidolph Instruments Labortechnik	Schwabach, DE
Ziegra Eis Machinen	?!, DE

2.2 Antibodies

Name	Species	Manufacturer
8oxoG	?	?
other oxStress	?	?
Fibronectin	?	?
secondary ones	?	?

2.3 Celllines

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

2.4 Primer

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

(Continued)

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

2.5 Chemicals

Name	Manufacturer
5X First Strand Buffer	Invitrogen™ (Marke von Thermo Fisher Scientific Inc.)
Antimycin A	Sigma-Aldrich Co. LLC.
BSA	?!?
CellROX	Thermo Fisher Scientific Inc.
Collagen Type I, rat tail	Ibidi -> loo this up
dNTP Mix	AB -> look this up
DTT	Invitrogen™ (Marke von Thermo Fisher Scientific Inc.)
Ethanol (concentration?)	J.T. Baker -> look up
FCCP	Sigma-Aldrich Co. LLC.
D-Glucose	Sigma-Aldrich Co. LLC.
GlutaMAX™-I	Gibco BRL
Hoechst	?!?
IL-1	Pepro Tech -> look this up
M-MLV RT	Thermo Fisher Scientific Inc.
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	Pepro Tech -> look this up
Sodium Pyruvate	Gibco BRL
RiboLock R	Thermo Fisher Scientific Inc.

(Continued)

Name	Manufacturer
Seahorse XF calibrant	Agilent Technologies, Inc.
PowerUp™SYBR™GREEN Master Mix	Thermo Fisher Scientific Inc.
TGFβ	Pepro Tech -> look this up

2.6 Media, Supplements

Name	Manufacturer
FBS Gold Plus	Biosell -> look this up
Medium 231 (M231)	Gibco BRL
Smooth Muscle Cell Growth Supplement	?!?
XF Base Medium	Agilent Technologies, Inc.

2.7 Solutions

Name	Manufacturer
Interleukin 1 beta (IL-1)	IL-1 0.1 % BSA in PBS
N-Acetylcystein (NAC)	0.25 M NAC in water, ~pH 7
Platelet-derived growth factor-BB (PDGF-BB)	PDGF-BB 0.1 % BSA in PBS
Transforming Growth Factor beta (TGF)	TGFβ 0.1 % BSA in PBS

2.8 Kits

Kit	Manufacturer
Total RNA Purification Kit	Jena Bioscience GmbH

2.9 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 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	Assistent - look this up
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	?!?
0,20 µm filter	?!?
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

2.10 Devices

Name	Manufacturer
Bench Heraus	Heraeus Holding GmbH
Bench 2	
Research pipettes (2.5µL, 10 µL, 100 µL, 1000µL)	Eppendorf AG
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 AG
Eclipse TS100	Nikon Corporation
pipet X rainin	?!?
Centrifuge 5415 R	Eppendorf AG
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 ?!
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	Thermo Fisher Scientific Inc.
Ice maschine	Ziegra Eis Machinen

2.11 Programs & Modules

Programs

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

(Continued)

Program	Version	Manufacturer
keyence sftware?!		
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 Im
sqlite3_analyzer	3.38.5.	The SQLite Consortium
Wave Controller	2.6.3	Agilent Technologies

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 have been written and kindly provided by Till Tantau.
– Styling of the GWAS Visualizer was done with the CSS Framework Bootstrap.

2.12   Public Data

3

Methods

3.1 Cultivation and differentiation of HAoSMCs

For the following experiments human arortic 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 when ever possibile.

Cells were differentiated treated first with Transforming Growth Factor beta (TGFβ) and then with Interleukin 1 beta (IL-1) & Platelet-derived growth factor-BB (PDGF-BB) to induce a synthetic phenotype. For more imformation please check the section 1.4 on smooth muscle cells in CAD.

Thawing & Cultivation

Cells were cultivated to a maximum passage of 10, after that new passage cells were thawed. For long time storage cells were kept in [cryo medium] and stored in liquid nitrogen. When required, new cells (6th passage) were need cells were thawed at 37°C in the water bath and transfered to a falcon. After centrifugation for 2 min at 300xg the supernatant was removed and the cell pellet taken up in 14 mL of M231 + Smooth Muscle Cell Growth Supplement (SMGS) for cultivation in a T75 flask. Every other day 2/3 of the medium were removed and replaced by fresh.

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). Then the cells were incubation with 3 mL trypsin for 4 min at 37°C. After 7 mL M231 were added to the deattched cells and the cells were transfered to a falcon and pelleted for 4 min at 300xg. The supernatant was removed and the pellet resuspended in M231 + SMGS, seeding 500×10^3 cells per T75 flask.

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 collagen 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 transfered in each used well of a 24-well plate, fully coating the bottom of the well. For polymerization the matrix was incubated at 37°C for at least 60 min.

Differentitation of HAoSMCs

Differentiation was carried out in 24 wells plates with 1 mL M231 supplemented with 1 % fetal bovine serum (FBS) and different cytokines:

- **Day 0:** Matrix and cells were prepareed as described in the sections Preparation of col I matrix and Passaging. Seeding of 40×10^3 in M231 + SMGS on plastic or on 160 µL col I matrix.
- **Day 1:** After 24 h the medium was replaced with 1 mL M231 + 1% FBS + 5 ng/mL TGFβ (or just 1 mL M231 + 1% FBS).
- **Day 5:** The medium was replaced with 1 mL M231 + 1% FBS + 10 ng/mL IL-1 + 10 ng/mL PDGF-BB (or just 1 mL M231 + 1% FBS).
- **Day 7:** Potentially further stimulation described in the section of the used assay.

3.2 mRNA Quantification

quantative polymerase chain reaction (qPCR) was utilized to assess the mribonucleic acid (RNA) concentration of the two reporter genes Calponin 1 (CNN1) and Matrix metallopeptidase 9 (MMP9) in HAoSMCs differentiated as described in section on differentiation. Using the house keeping gene Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for reference.

SYBR® Green is an intercalating DNA dye that allows for the monitoring of DNA amplification. Flourescence is measured after every amplification cycle of the PCR yielding a crossing point when signal reaches a certain threshhold. A lower quantification cycle (Cq) corresponds to an higher inital DNA concentration (**huggettStandardisationReportingNucleic2011**).

RNA Isolation

RNA was isolated using the kit and extraction was performed according to the corresponding protocol, using an extra washing step with 700 µL 80 % ethanol and eluting with 30 µL of RNase-free water. Determination of nucleic acid concentration was carried out with the NanoDrop.

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 described in the following table:

Table 3.2: RT Master Mix

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

The reaction 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

The samples were prepared in a 384-well plate using SYBR® Green Master Mix:

Table 3.3: qPCR samples The samples for the 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

Wells were sealed, thoroughly mixed by invertation of the plate and the assay performed using follwing programme on the TaqMan:

Table 3.4: qPCR programme The programme for the qPCR

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 ct of the reference gene GAPDH was substracted from the mean ct 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 an 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 was done in python using the modules: pandas, numpy, scipy as well as pyplot and seaborn. Assuming a normal distribution, student's t-test was used, a p-value of 0.05 is considered as significant. For detailed information please check the script.

3.3 Energy Profiling

Seahorse Assay was utilized to assess the energy profile of HAoSMC differentiated as described in section on differentiation. For this assay cells were not differentiated in a 24 well plate but the seahorse plate, using 5 technical repeats and on control well for 4 tested conditions. Since the plate would not fit the matrix cells were cultivated in plastic!
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 glycolisis respectively and can be used to assess the metabolic function of cells (**HowAgilentSeahorse**).

Seahorse Assay

On the day before the assay the Seahorse XF Analyzer was turned on to calibrate and the sensor cartrigae was left to equilibrate in Seahorse XF calibrant over night at 37 °C (in non-CO2 environment).
On the day of the assay, cells were washed with 500 µL PBS each and afterwards 500 µL supplemented XF BASE medium were added, cells were left to incubate for 1 h at 37°C in non-CO2 environment. During this time toxins for disruption of the respiratory chain were prepared and added to the catridge:

Table 3.5: toxins for seahorsetoxins for seahorse. :)

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 compound catridge was loaded into the XF Analyser for calibration, after successful calibration the hydration catridge was replaced with the cell plate. Measurement was carried out as following:

– Calibration of the probes.

- Equilibration
- 3 Repeats of:

    - Mixing (1 min)
    - Pause (2 min)
    - Detection of OCR and EACR (4 min)

- Pause (2 min)
- Injection of 55 µL Oligomycin
- 3 Repeats of:

    - Mixing (1 min)
    - Pause (2 min)
    - Detection of OCR and EACR (4 min)

- Pause (2 min)
- Injection of 60 µL FCCP

    - Mixing (1 min)
    - Pause (2 min)
    - Detection of OCR and EACR (4 min)

- Pause (2 min)
- Injection of 55 µL Antimycin
- 3 Repeats of:

    - Mixing (1 min)
    - Pause (2 min)
    - Detection of OCR and EACR (4 min)

Finally the medium was removed and cells were stained for 15 min with Hoechst (concentration in PBS) and photographed in the keyence to determine cell count for normalization.

Processing of Data

Cells were quantified using (what exactly does Tobias script do) with a python script provided by my supervisor Dr. Tobias Reinberger, cell count and the signal of the control wells were used to normalize the OCR and EACR calculated by the XF Analyzer with the accompanying software.
In total three biological were recorded of which on was excluded because no changes in OCR and EACR could be detected and cells deacttched from the bottom of the wells during Hoechst staining. For the remaining two replicates the least fitting of the 5 technical repeats for each condition was manually excluded.
Further , again using a modified python script provided by Dr. Tobias Reinberger. Assuming a normal distribution, student's t-test was used, a p-value of 0.05 is considered as significant. For detailed information please check the script.

3.4 Oxidative Stress Assay

CellROX Green assay was used to assess generation of reactive oxygen species in HAoSMC differentiated as described in section on differentiation, after further stimulation (from here

on refered to as 'boost') with PDGF. Additionally a recovery experiment was performed using NAC, a potent anioxidant, to quench generation of ROS.
CellROX Green is a flourescent dye that gets oxidized by ROS and then binds to DNA, showing bright green flourescence (**CellROXGreenReagent**).

CellROX Assay

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

Table 3.6: Seahorse Assay

component	concentration	final concentration	volume (µL)
HBSS	-	-	300
PDGF	?	variable (0 - 400 ng/mL)	variable
Hoechst	?	?	0.3
CellROX (1:500)	?	?	0.6
NAC	0.25 M	variable (0 - 8 mM)	variable
total	-	-	$\sim$300

Cells further kept at 37°C in 5 % CO2 environment, incubation time is indicated with the results of the respective experiment. Imaging was done at the keyence microscope using standard sensitivity and a exposure time of s in the green channel and s for the blue channel.

Processing of Data

For time resolved PDGF-BB boost titration 7 biological repeats were performed of which one was excluded because of high signal in the negative control. For NAC quench 4 biological repeats were performed of which one was excluded because no signal in the positive control. For quantification of signal itensity, pixels with a green value higher than 90 were counted. Differents in cell count were adjusted by divison by the number of pixels with a blue value bigger than 80 (CHECK TRESHHOLD VALUES!!). Z-STACK FOR THE NAC quench images. To adjust for large variance in total signal intensity between biological repeats, values were adjusted by divison through the total signal of all recorded conditions.
Mann Whitney U Test was used, a p-value of 0.05 is considered as significant. For detailed information please check the scripts.

3.5 Immunoflourescence

Fibronektion as marker of matrix. Used cells. Maybe also the anti-8-oxoG AB?

Protocol

First cells were washed with PBS and fixated for 40 min with 200 µL -20°C methanol-aceton (1:1), after the removal of methanol-aceton cells were left to dry for 20 min. Then cells were

treated for 30 min with 250 µL permeabilization buffer, folllowed by 30 min with 250 µL blocking buffer and incubated with 300 µL of the primary AB over night at 4°C:

The next day the primary AB was removed and cells wre incubated for 60 min at RT with the secondary AB:

After removal of the secondary AB cells were washed 3 times with PBS and stained for 15 min with DAPI (1:5000 in water).

Processing of Data

Me counting pixels.

3.6   Curation of Data for postGWAS Analyses

GWAS data and data for postGWAS analyses and co-visualization was downloaded from public resources. Processing of the data and further annotation is briefly described in the follwing listing. For a complete overview of all the data collected please refer to table or the download scripts. For explainaition of the data please refer to the introdcution or the sources themselves.

- **GWAS Summary Statistics:** The CAD GWAS summary statistics as well as a list of indetified proxy SNPs from the study was annotated with [what did Tobias do] via the ensembl REST API by Dr. Tobias Reinberger.
- **HGNC Gene List** The newest quarterly update to the complete HGNC dataset was downloaded via the FTP server. The dataset was used to generate a list of all approved symbols, mapping to their HGNC id. Further a list of all symbols (approved, alias and previous) was generated, 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). In the same process linked SNPs were annotated with their most severe consequence (by VEP) via the ensembl REST API.

Table 3.7: 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 FTP server. Features annotated as genes of the type protein coding, lncRNA or miRNA were extracted, further gene symbols were mapped to their HGNC id if possible.
- **Open Target Genetics l2g Scores** The lattest list of Open Target Genetics l2g Scores was downloaded via the FTP server. Entrie were annotated with their HGNC

ID when ever possible, entries that do not map to a gene that is approved by the HGNC were dropped.

- **Ensembl Regulatory Build** The newest ensembl regulatory build (ensembl release 106) was downloaded via the FTP server.
- **TSS** Transcription start sides for protein coding genes were extracted from a USCS Brower dump.
- **Associated traits from GWAS catalog** The SNP trait associations from the lattest release of the GWAS catalog as well as the accompanying list of studies was downloaded via the FTP server. SNP-trait correlations missing a the position or a p-value for the association were dropped from the data set. Further the column for Odds Ration or beta was seperated in to two columns.
- **TADs** The data on TAD was downloaded via the download link on the 3D genome browser.
- **scATAC-seq from Miller et al.** The processed scATAC seq data was scraped from the Miller Lab GitHub repository.
- **scATAC-seq from CATlas** The processed scATAC seq data was scraped from the Ren Lab website.
- **ABC model** The ABC model data was downloaded from the Engreitz Lab FTP server. The data was translated from Hg19 to Hg38 with pyliftover.
- **ENCODE cCREs** Done by Dr. Tobias Reinberger.

3.7 Visualization of GWAS data

For visualization of the data, an bokeh application was build, that fetches the data from the database and renders it to a webbrowser.

Bokeh is a python module that allow 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 JSON data and handed over to bokehJS which deserializes them into JavaScript objects that are rendered to the browser. The intergrated bokeh server additionally offers the possiblity to synchronize data between the underlying python environment and brower 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 sections. Reading of data from the database and further processing steps are managed by a data provide and enclosed in one class. In contrast to the model-controler-view architecture, a popular architectural pattern for the design of user interfaces, there is no partition between a view and a controler. Since data visualization as well as the control widgets are created by bokeh, it is convienient to use the build in event listeners of the library to handle the required callbacks. Therefore the main file is responsilbe for the creation of all plots and widgets as well as listening for inputs.
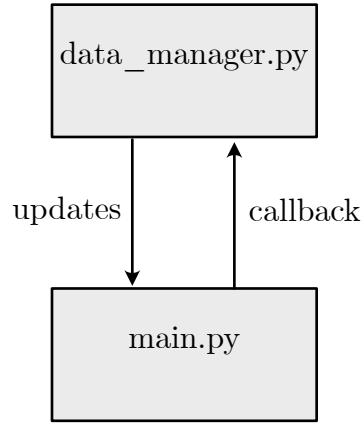
STYLING USING A bootstrap thingy

Figure 3.8: Architecture of the vis tool

3.8   Enrichment analysis

Based in 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-represent in a list of interest. The most prominent application probably being their application as gene set enrichment analyses (GESA). GESA are used to check for the over-representtion of a candidate gene list in a predefined set of genes (**tipneyIntroductionEffectiveUse2010**). In this case the method is used to determine if regulatory elements which overlap with SNPs associated with CAD are enriched in a biosample, using fisher's exact test.
As a list of regulatory elements the cell type specific cCREs that are part of the ENCODE project were used (excluding cCREs marked as unclassified). As a list of SNPs the list of CAD associated proxy SNPS and linked variants (european population, r2 >= 0.6) was used. The four values required for calculation of the enrichment factor () and p-value require are shown in the Contingency of figure:

- Number of distinct cCREs among all biosamples (m)
- Number of distinct cCREs that are annotated in the biosample of interest (mt)
- Number of distinct cCREs that overlap with a SNP in the SNP list in any biosample (n)
- Number of distinct cCREs that overlap with a 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 ($\equiv$ number of biosamples):
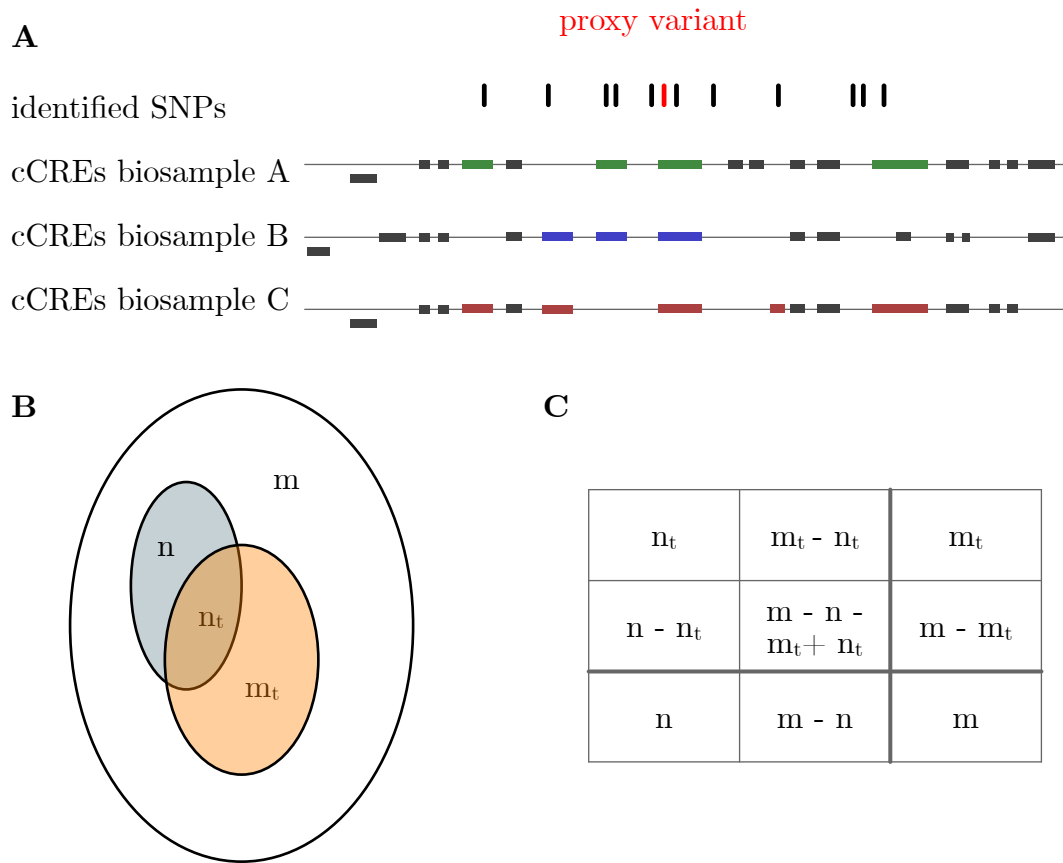
$$p_{ajd.} = p * n$$

Figure 3.9: enrichment

The analysis and visualization were done in python. A p-value of 0.05 is considered as significant. For detailed information please check the analysis script and the visualization script.

4

Results

4.1 Differentitaion

To generate a controlable phenotype, that is representive for the synthetic phenotype in vivo, HAoSMC were first treated for two days with TGFβ to induce a phenotype that ressambles the contrat phenotype. Using this contractile phenotype as a precursors, cells were further treated for four days with IL-1 and PDGF-BB which should induce differentiation into the synthetic phenotype.

Expression of CNN1 & MMP9

To track the differentiation and confirm the protocoll, the HAoSMCs were characterized using qPCR, assessing mRNA levels of CNN1 as a contractivle marker and MMP9 as a synthetic marker. For better compariability mRNA levels are considered in realtion to the house keeping gene GAPDH.

As seen in figure 4.1, stimulation of HAoSMCs cultivated on a col I-Matrix with TGFβ causes a significant increase in CNN1 expression, while not significant with four repeats, expression of CNN1 seems to drop off after further stimulation with IL-1 & PDGF-BB. The same trend, even if not as pronounced and not significant with four biological repeats, can be observed for HAoSMCs cultivated on plastic. Expression of MMP9 is not significantly different between any of the tested conditions, still a trend can be observed, that cells cultivated on the col I-matrix seem to show higher expression of MMP9.

Energy profile

Further the energy profiles of the cells were assesed via Seahorse Assay. oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) for HAoSMCs differentiated under the same conditions as for qPCR are displayed in figure 4.2.

It is important to note, that the assay was carried out on plastic because the matrix does not fit into the confined compartment created by the pistion for detection of OCR & ECAR. The OCR show an expected drop after inhibition of the ATP synthase with with Oligomycin, they further show a great increase after decoupling with Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP) and a steep drop-off after inhibition of coenzyme Q-cytochrome c reductase (complex III) with Antimycin A. Characteristics of the respira-
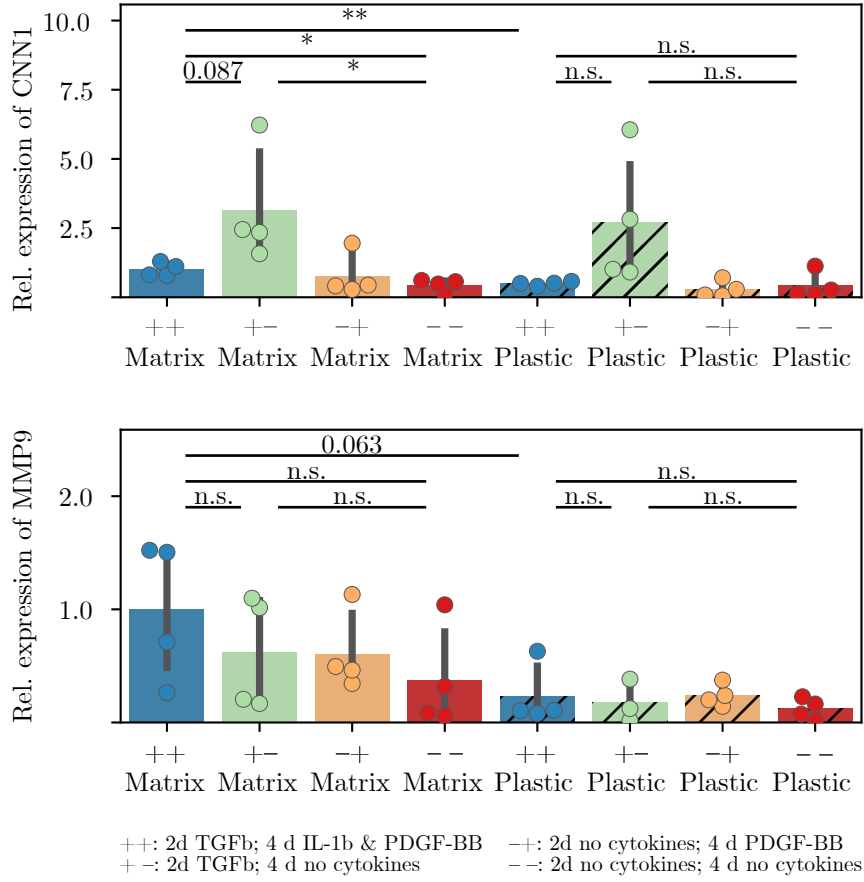
Figure 4.1: Relative Expression of CNN1 & MMP9 in HAoSMCs
qPCR analysis of expression for contractile marker CNN1 (top) and synthetic marker MMP9 (bottom) for HAoSMCs differentiated with different combinations of cytokines: **++:** 2 d with TGFβ followed by 4 d with IL-1 & PDGF-BB; **+−:** 2 d with TGFβ followed by 4 d without stimulation; **−+:** 2 d without stimulation followed by 4 d with IL-1 & PDGF-BB; **—:** 6 d without stimulation. All four conditions were tested on two differen surfaces (plastic vs. collagen 1 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: $* : p < 0.05; ** : p < 0.01$

tory chain were calculated as described in section 3.3. Looking at the energy profile of the cells 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 IL-1 & PDGF-BB (fig. 4.3, A).

More interesting difference can be observed when examining characterics of the respiratory chain. Stimulation with only TGFβ causes a sigbnificant decrease in basal respiration, ATP production, maximal respiration as well as spare capacity (figure 4.3, B top). Further stimulation with IL-1 & PDGF-BB then causes again significant increase of these paramters to similar levels as in undifferentiated HAoSMCs. Finally, it needs to be considered that

these observations are only based on two biological repeats done in plastic! To draw solid conclusions from seahorse assay, more biological repeats would be required.

4.2   Evaluation of oxidative Stress

It was evaluated if further stimulation of the induced synthetic phenotype with higher concentrations of PDGF-BB would yield ROS generation to an extend that can't be compensated by the ROS defense.

PDGF boost of out cells indcues oxidative stress

For this one repeats of an experiment already previously done in the group was carried out. Stimulating the four tested conditions for two hours with 200 ng/mL PDGF-BB in HBSS.

As displayed in figure 4.4 only stimulation for two days with TGFβ, followed by 2 days with IL-1 & PDGF-BB, followed by 2 h boost with PDGF-BB was able to trigger noticable generation of reactive oxygen species (ROS).

Characterization of the CellROX Assay

To get a better understanding of the assay and its limits a titration was carried out. For this HAoSMCs differentiated into the synthetic phenotype were boosted with different concentrations of PDGF-BB and CellROX signal was detected after 60, 120, 180 & 240 minutes.

As is clear from figure 4.5, a large impact factor on signal intensity is PDGF-BB concentration during the boost. Over all biological repeats only negliable signal was visible for concentrations < 100 ng/mL, followed by a significant increase in signal at 200 ng/mL. Signal intesity for boost with 400 ng/mL was mixed, in some cases yieling even higher signal, in some cases collapsing. The second large influence that is clearly resolved in this assay is duration of the boost. After 120 min first signal is obserable, intesifying at 180 min and 240 mins. At 240 min also medium signal can be oberved for all the conditions, indicating an unspecific response.

While the trend within on biological repeat was consistent and reproducable, variance between repeats was almost to the same extent as differences between the conditions. Two potential causes that are out of control of the experimentor are degradation possible degradation of CellROX Green which should be used up within 6 month after delivery as well as the exact composition of the col I matrix since it was obeserved that intensity in the CellROX assay seems to correlate with col I concentration in the matrix (data not shown).

In an attempt at normalization (figure 4.6), trying for account for these external factors, all data points were selected that were used in all biological repeats and itensities were normalized by the cummulative intensity of all selected conditins. This is quite similar to normaliazing to the highest signal observed for the biological repeat because signal at 200 ng/mL after 240 min has by far the biggest contribution to the cummulative value of the biological repeat.

Rescue of ROS production using NAC

Finally, trusting the trend over all trend of the titration, a rescue experiment was performed, to verify that observed signal in the CellROX assay was indeed due to generation of ROS.

For this ROS generation was quenched by the addition of 2, 4, or 8 mM of N-acetylcystein (NAC) Quenching the signal with NAC.

Confirmation that the signal is indeed due to generation of intracellular ROS.

Finally it should be noted, that the signal onyl build up over 15 - 20 minutes under the microscope after the cells were taken out of the incubator. This indicates that generateion of ROS might not exclusively triggered by PDGF-BB stimulation but also required additional contributors like the loss 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 anyway to image all wells.

4.3   Database and GWAS Visualizer

This was quite a long process, tinkering around with different designs to get a working solution. For this use case visualization in the browser using a database as the backend was the final decision. Compare to fetching everything online -> slow and just grabbing everything from files -> extremely stressful to maintain.

Curation of Data

Describe what actually happend to the data, probably in the methods section. This should be pretty self explanatory. Just have a look at the table and figure.

Visualization

Building on the initially intended use case for the data a visualization tool was build as briefly described in the methods section, for more details please check the source code or just ask.

Describe what the tool can do. Search function for SNPs, also for genes, providing info for associated SNPs. Visualization of the window SNPs, including linked SNPs annotated with important info such as most severe consequence. Further the data is aligned with important stuff like overlapping genes (integrating l2g scores). Also enesembl regulatory build and TADs. Further a lot of cell specific data in the form of scATAC seq tracks and promotor gene links from the ABC model. These are also shown and for easier navigation grouped into different classes using cellosaurus. To have a better look a individual variants these can be clicked, only highlighting tracks that overlap with the specific selected variant.

For more information regarding the aligned data please check the corresping paragraph in the intrdocution.

4.4   Enrichment analysis

The only data that is not displayed in the plot are cCRE elements which were used for an enrichment analysis. Checking different biosamples for significant enrichment of candidate cis regulatory elementss (cCREs) that are overlapping with proxy SNPs identified in the CAD GWAS or variants that are in LD with these. The procedure is described described in methods.

As seen in figure 4.9, statistical signifcant enrichment ($p_{adj.} < 0.05$) was observed for 34 biosamples. These biosamples were annoted using Cellosaurus data. The most prominent groups of associated tissues were heart (8), lung (7) and artery (6).
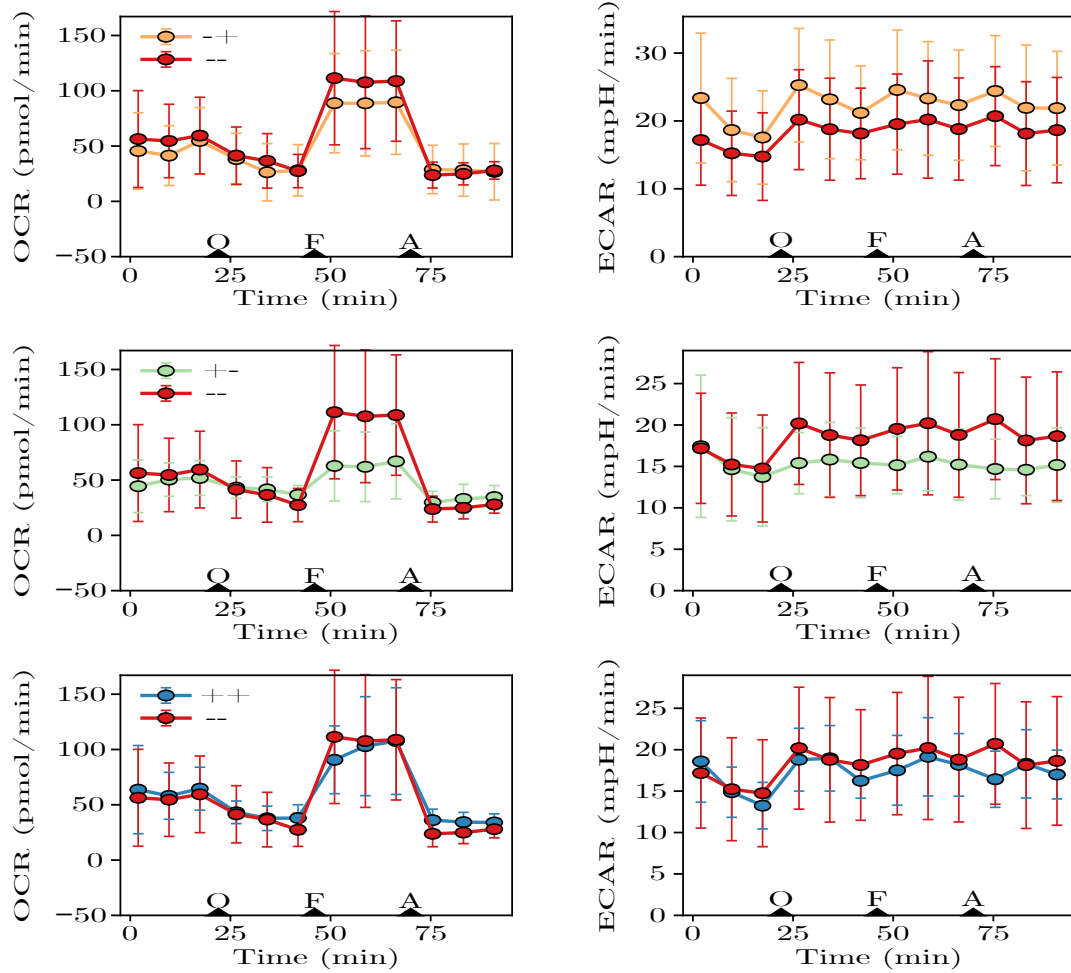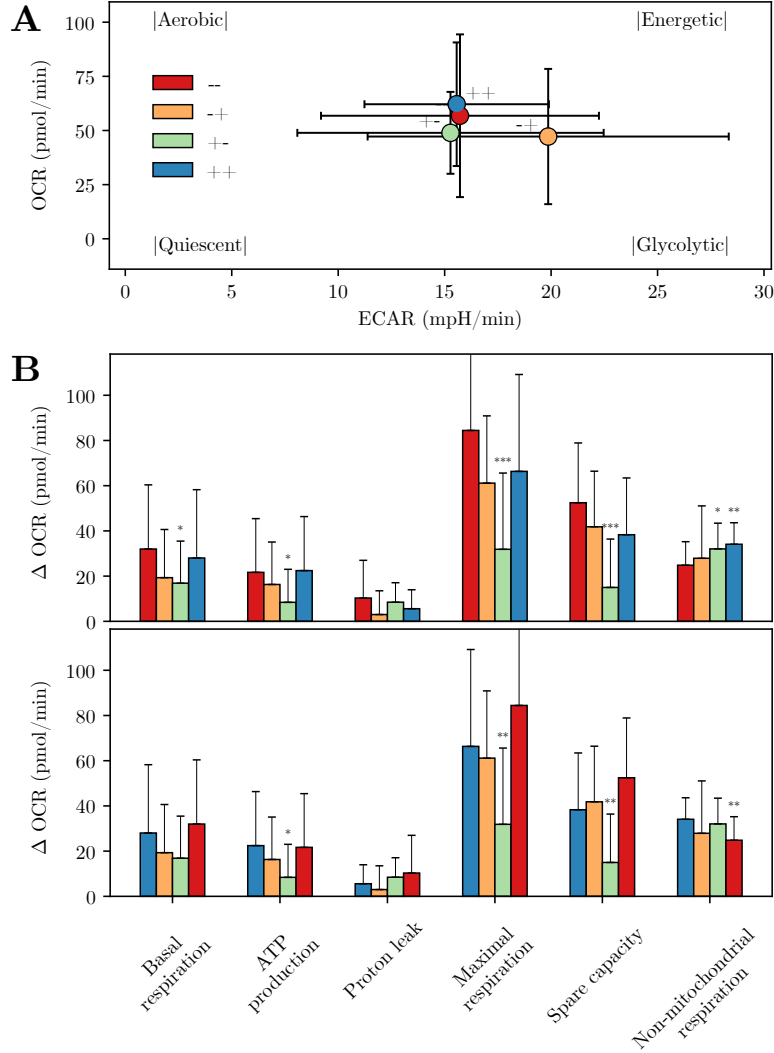
Figure 4.2: OCR & ECAR of HAoSMCs

Seahorse assay for HAoSMCs differentiated with different combinations of cytokines. **++:** 2 d with TGFβ followed by 4 d with IL-1 & PDGF-BB; **+−:** 2 d with TGFβ followed by 4 d without stimulation; **−+:** 2 d without stimulation followed by 4 d with IL-1 & 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.

Figure 4.3: Energy profile of HAoSMCs
Seahorse assay for HAoSMCs differentiated with different combinations of cytokines. $++$: 2 d with TGFβ followed by 4 d with IL-1 & PDGF-BB; $+-$: 2 d with TGFβ followed by 4 d without stimulation; $-+$: 2 d without stimulation followed by 4 d with IL-1 & PDGF-BB; $--$: 6 d without stimulation. (**A**) Initial OCR & ECAR of the four tested conditions. (**B**) Characteristics of the 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: $* : p < 0.05; ** : p < 0.01, ; *** : p < 0.001$
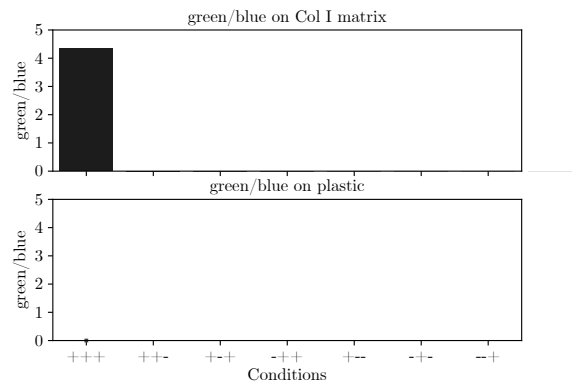
Figure 4.4: Stimulation with PDGF induces oxidative stress.
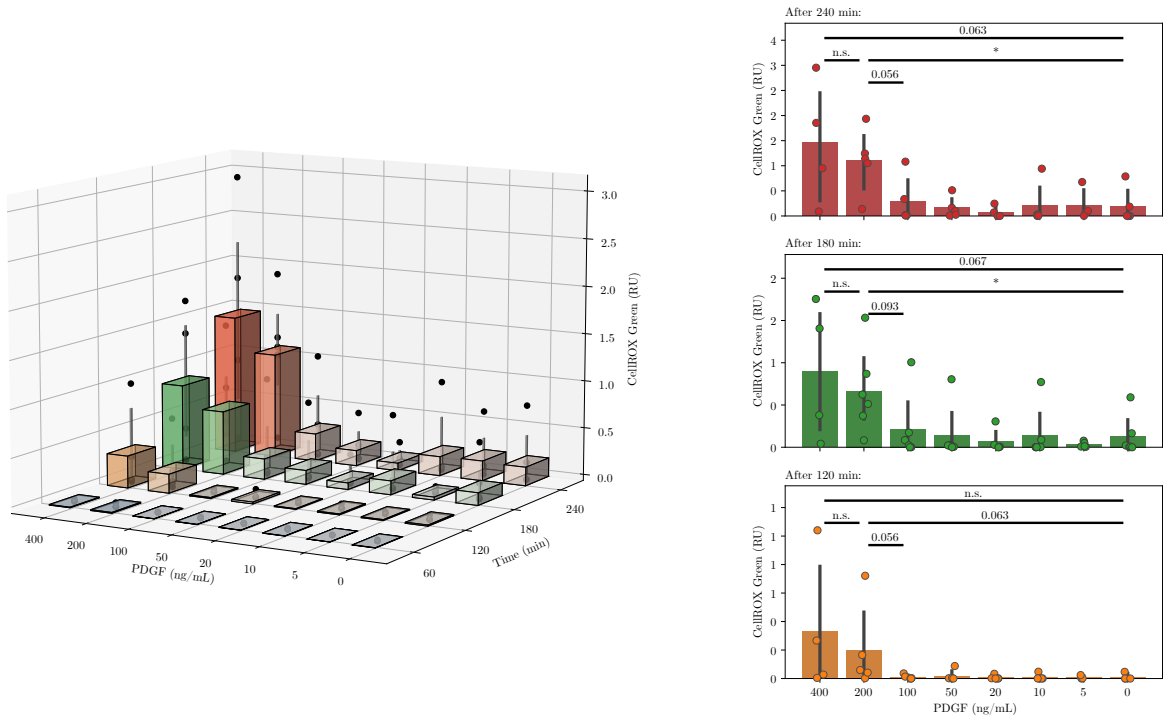Repeat of the result already shown by Tobi.
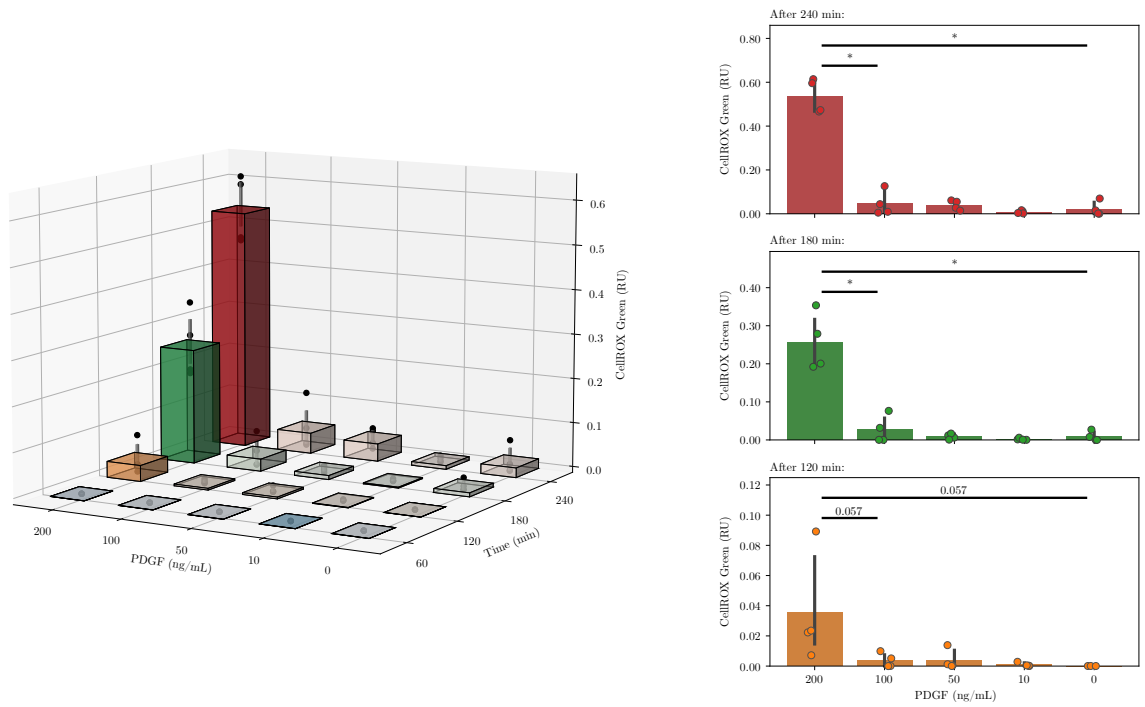

Figure 4.5: CellROX titration Titration

Figure 4.6: Stimulation with PDGF induces oxidative stress - normalized.
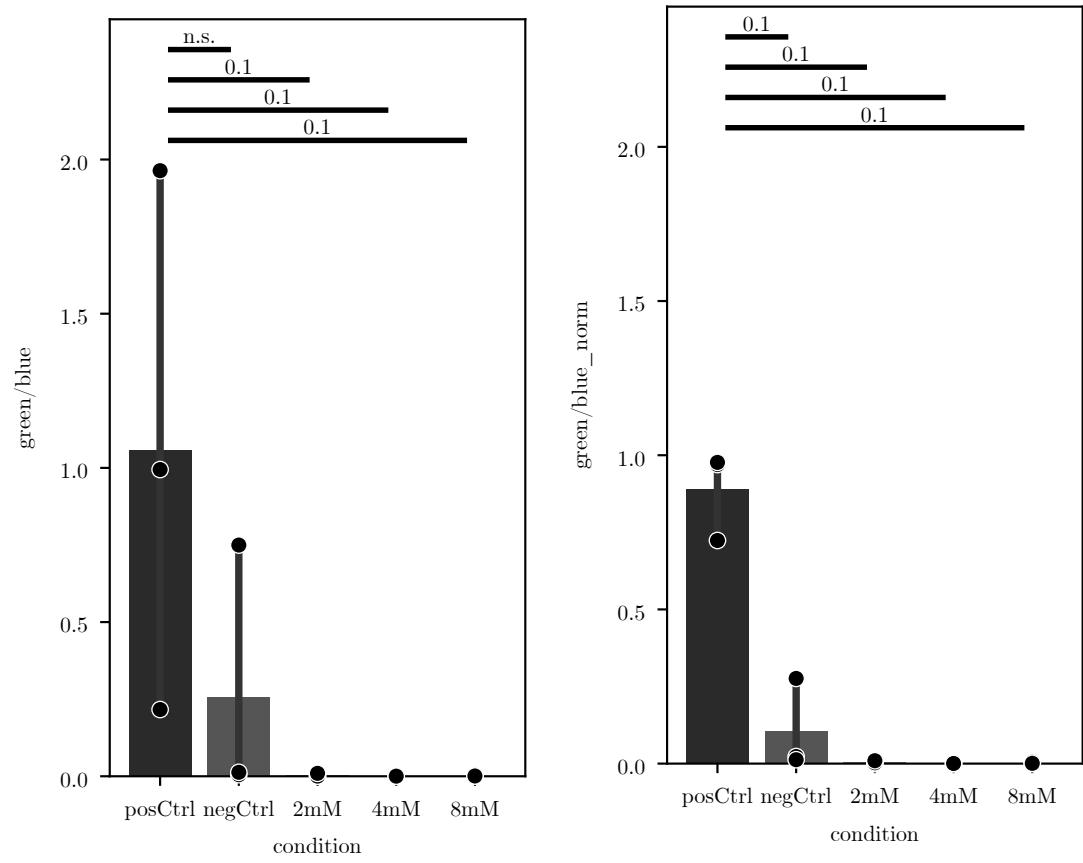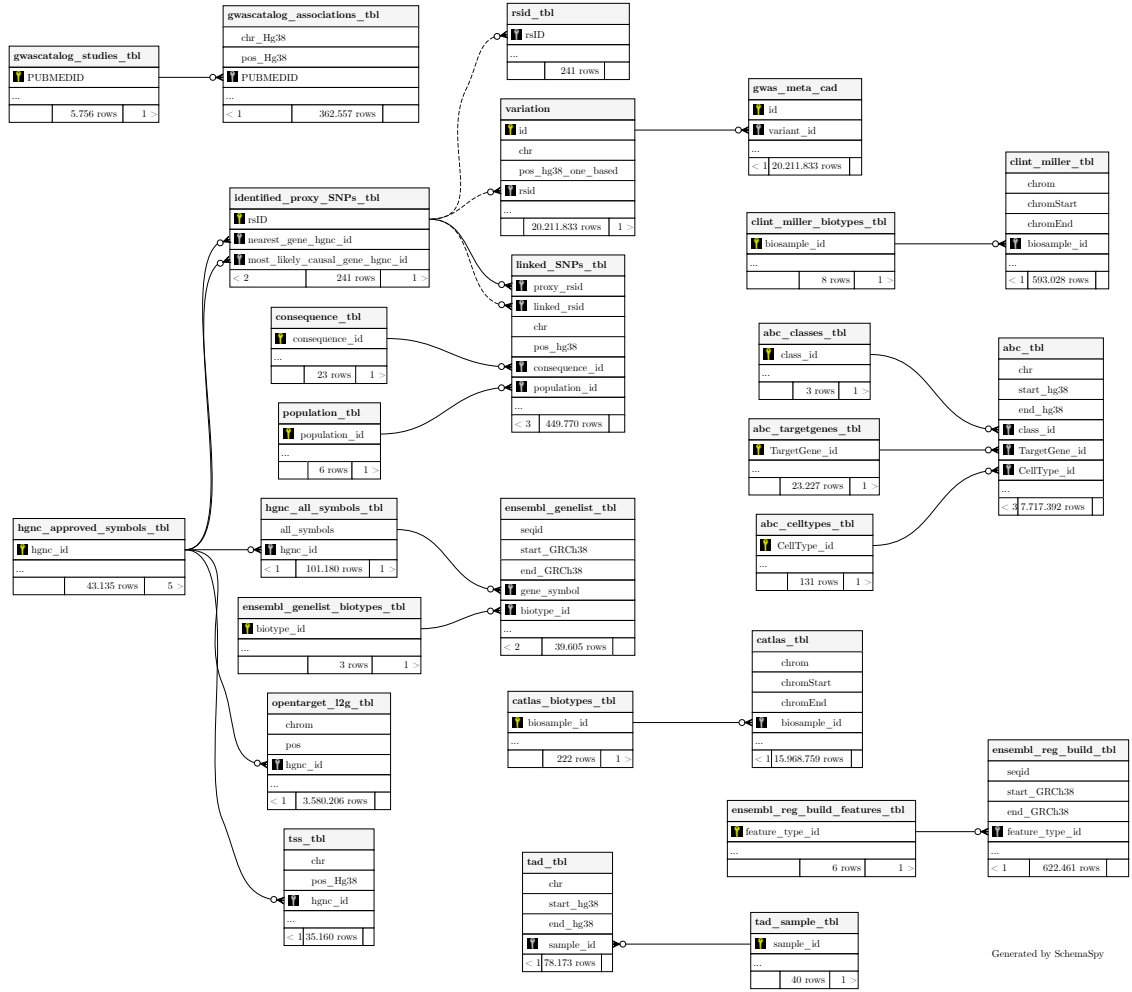
My attempt at normalization.

Figure 4.7: NAC quench

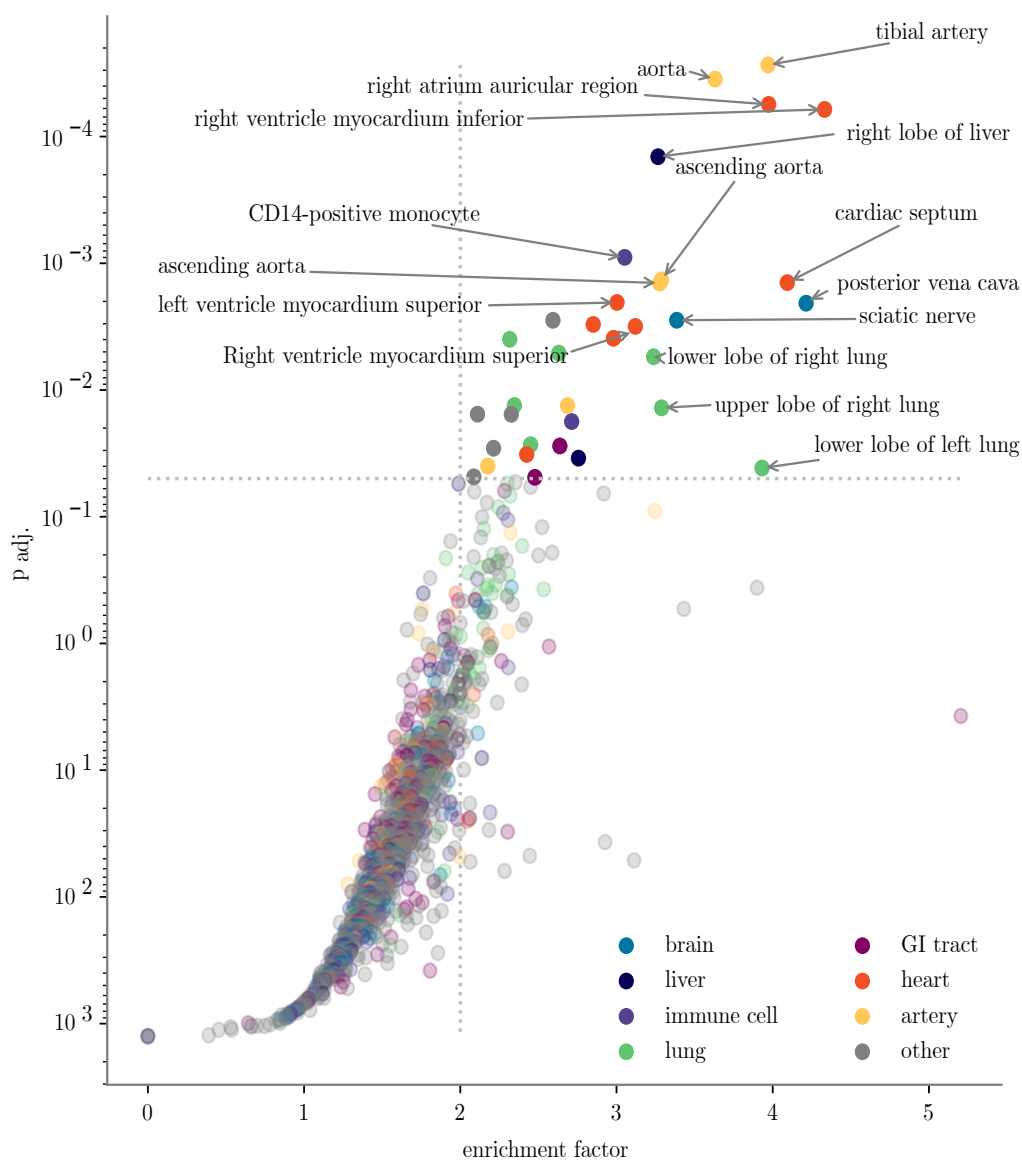**Figure 4.8: Entity-Relationship Diagram of the Database**

Figure 4.9: Enrichment Analysis This stuff actually seems to be working!!

Table 4.10: Enriched tissues

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

Here I'll discuss my results should I ever finish the rest of my thesis.

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.