BART DEPLANCKE

PERSONAL INFORMATION

Bart DEPLANCKE, PhD; Date of birth: 21/08/1975; Nationality: Belgian; Web site: http://deplanckelab.epfl.ch/ Researcher unique identifier(s) (ORCID): 0000-0001-9935-843X

EDUCATION

2002	Ph.D. in Immunobiology (Division of Nutritional Sciences), University of Illinois, USA
1998	M.S. in Biochemical Engineering, Ghent University, Belgium
1995	B.S. in Bio-engineering, Ghent University, Belgium

EMPLOYMENT HISTORY

2020–present	Full Professor in Systems Biology and Genetics, Institute of Bioengineering, SV, EPFL
2018-present	Vice-Dean of Innovation at the School of Life Sciences (SV), Ecole Polytechnique Féderale de Lausanne (EPFL), Switzerland
2013–present	Swiss Institute of Bioinformatics Group Leader
2014–2019	Associate Professor in Systems Biology and Genetics, Institute of Bioengineering, SV, EPFL
2007–2014	Assistant Professor in Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, EPFL (CH)
2003–2007	Postdoctoral Fellow, Program in Gene Function and Expression, Program in Molecular Medicine, University of Massachusetts Medical School, Worcester (USA)

2002–2003	Postdoctoral Fellow, Department of Cancer Biology in Dana-Farber Cancer Institute & Department of Genetics, Harvard Medical School, Boston (USA)
1998–2002	Research Fellow, Division of Nutritional Sciences, University of Illinois (USA)
1997–1998	Research Assistant, Department of Biochemical & Microbial Technology, Ghent University (BE)

INSTITUTIONAL RESPONSIBILITIES (a selection)

Vice-Dean of Innovation at the School of Life Sciences (since 2018)

EPFL Innovation Council (INC) (since 2018)

Faculty Direction Member (since 2018)

Member of the **Doctoral School in Quantitative and Computaional Biology Reviewing Committee** (since 2018)

Main Coordinator of the **Catalyse4Life EPFL SV Research Innovation Program** (since 2017)

EPFL Life Sciences SV-IT Steering Committee Member (2017–present)

Main Coordinator of the 2016 EPFL School of **Life Sciences Audit**

SUPERVISION OF > 20 GRADUATE STUDENTS AND 15 POSTDOCTORAL FELLOWS (since 2008)

TEACHING ACTIVITIES (summary)

Genetics & Genomics (4 credits; 7x3 h; 3rd year Bachelor); Single Cell Biology (4 credits; 7x3 h; 1st year Master)

GOVERNING ACTIVITIES

Search Committee Member for candidates of Assistant Professorships at the EPFL (Bioinformatics, Bio- engineering, BioMEMS, BioPhotonics, Biostatistics, Cancer, Informatics & Communication, Metabolism, Microbiology, Statistics); INSERM ATP-Avenir (France; Genetics & Genomics); KU Leuven (Belgium); Institute of Technology (Italy; Systems and synthetic biology); ETH Zurich (Systems Biology & Bioinformatics); University of Lausanne (Computational Biology; Switzerland) and University of Fribourg (Environmental Biology / Omics)

Ad-hoc paper reviewer (since 2006, a selection): Cell, Development, eLife, Genome Biology, Genome Research, Molecular Cell, Molecular Systems Biology, Nature, Nature Biotechnology, Nature Cell Biology, Nature Communications, Nature Genetics, Nature Methods, Nature Medicine, Nature Reviews Genetics, Science

Editorial Activities: Editorial Board Member of BMC Genomics and Nucleic Acids Research; 2017 Issue of Current Opinion in genetics and Development (Genome architecture and Expression topic)

Member: 1) Center for Organismal Studies Scientific Advisory Board, University of Heidelberg (from 2019); 2) Health 2030 Genome Center Strategic board, Geneva, CH (from 2018); 3) "Integrative Biology of the Cell" Research Center Scientific Advisory Board, Paris-Saclay University (from 2018); 4) Swiss National Science Foundation National Research Council (from 2017); 5) Lausanne Integrative Metabolism Network Association Board of Directors (from 2017); 6) The ETH Domain Strategic Focus Area "Personalized Health and Related Technologies" Executive Committee (from 2017)

Reviewer for Grant Proposals and Programs: 1) Member of the Swiss National Science Foundation National Research Council, 2017-present; 2) Member of the Advanced Postdoc Mobility fellowship Evaluation commission of the Swiss National Science Foundation (SNSF, since 2014); 3) Reviewer of Starting, Advanced and Synergy ERC grants (since 2016); 4) National Infrastructures in Health and Biology Reviewer for the Investments for the Future (ANR) Program (France, 2016); 5) Functional Genomics Center Zürich (2014), 6) Swedish Foundation for Strategic Research (www.stratresearch.se); 7) FWO grants and postdoctoral fellowships (Belgium); 8) SNSF-based proposals: Div III (Biology and Medicine), Interdisciplinary Research, Sinergia, Ambizione

ORGANISATION OF SCIENTIFIC MEETINGS

1) Organizing Committee Member of the 2019 European Drosophila Research Conference, Lausanne (CH), September 2019; 2) Organizing Committee Member of the First (Leuven (Belgium), December 2017) and Second (Janelia Research Campus (USA), March 2019) Fly Cell Atlas meetings; 3) Organizing Committee Chairman for the 2017 International SystemsX.ch Conference (Zürich, CH); 4) Organizing Committee member of the 2010, 2015 & 2016 EPFL International Life Science Symposia; 5) Co-organizer (with Prof. Jeff Jensen, EPFL) of the Conference "Systems Genetics and Evolution of Non-human (Model) Organisms" (>100 participants), Ascona, CH (2014)

INVITED PRESENTATIONS (> 100 at International Symposia or Institutes, > 60 in the last 5 years)

FELLOWSHIPS AND AWARDS

2021 Cloëtta Prize for outstanding contributions to biomedical research.

Highest Level SNSF Project Grant Rating with invitation to **Bonus of Excellence** Program (2018)

Elected to the **National Research Council (2017) of the Swiss National Science Foundation**

EPFL Teaching Ambition Award (2012) for dedication to undergraduate teaching

Peter Reeds Young Investigator Award for 2005 by the American Society for Nutritional Sciences;

Henri Benedictus-BAEF Fellow of the King Baudouin Foundation and the Belgian American Educational Foundation in Biomedical Engineering, 2002–2003

College of Agricultural, Consumer and Environmental Sciences Doctoral Student Research Award for best PhD Thesis, University of Illinois, 2001.

CURRENT FUNDING ID

2020–2024 – Principal PI – SNSF Project Grant – Uncovering novel molecular principles underlying regulatory variation using variable chromatin modules;

2020–2021 – Principal PI – EPFL Open Science Fund – ASAP: an open, robust and interactive web-based portal for (single cell) omics analyses;

2019–2021 - Collaboration (2 groups, PI: Christian Wolfrum) – Precision Health and related Technologies (PHRT) Pioneer Grant - Adipose tissue heterogeneity and function in the development of metabolic diseases;

2020–2024 – H2020 – Innovative Training Network – ENHPATHY: Molecular basis of human enhanceropathies

2019–2023 – Consortium (5 groups, PI: Bart Deplancke) – Swiss National Science Foundation (SNSF) Sinergia – Elucidating the human mesenchymal bone marrow stromal hierarchy in health and disease;

2019–021 – Consortium (4 groups, PI: Christian Wolfrum) – PHRT Project Grant – Targeting the brown fat: Personalized strategies for treatment of metabolism;

2019–2022 – Principal PI – SNSF Project Grant – Dissecting the molecular and physiological function of Aregs in white adipose biology.

PUBLICATION RECORD

As of August 2021: h-index: 51; >10.7k citations (Source: Google Scholar); 114 total publications 88 research papers (36 as last author), 9 reviews, 3 perspectives, 6 protocols and 8 consortium-based.

Complete List: https://scholar.google.ch/citations?user=EMV2SUMAAAAJ&hl=en

TRANSLATIONAL ACTIVITIES

Co-founder Alithea Genomics (2020): a company specializing in high-throughput transcriptomics solutions for drug screening and biobank functionalization. Winner of the 2021 Swiss-wide VentureKick competition.

Co-founder and Former Chairman and Board Member of Genohm SA (2011): a company delivering software (big data management) products and services for pharma and R&D labs (www.genohm.com). Acquired by Agilent Technologies in May 2018.

Patents:

Patent application entitled **"High Throughput One-Hybrid System"**. Inventors: University of Massachusetts; Dana Farber Cancer Institute. Inventors: A.J.M. Walhout, M. Vidal, **B. Deplancke**. International publication number: **WO2005/005960**.

Patent application entitled **"Microfluidic device and method for isolation of nucleic acids"**. Inventors: A. Isakova, **B. Deplancke**. Applicant: Ecole Polytechnique Fédérale de Lausanne (EPFL). International publication number: **WO2016/059619**.

Patent application entitled **"Soft microbotic device for high throughput single cell studies"**. Inventors: Johannes Bues, Riccardo Dainese, Marjan Biocanin, **B. Deplancke**. Applicant: Ecole Polytechnique Fédérale de Lausanne (EPFL). International publication number: **WO2018/051242**.

Patent application (provisonal) entitled: **"A microfluidic device for rapid, multiplexed, bead-less, chromatin immunoprecipitation with on-chip DNA processing"**. Inventors: Riccardo Dainese, **B. Deplancke**. Applicant: Ecole Polytechnique Fédérale de Lausanne (EPFL). Application number: **PCT/IB2017/057889 (2017)**.

SELECTED PUBLICATIONS

(*, first author; #, corresponding)

M. Litovchenko, A.C.A Meireles-Filho, M.V. Frochaux, R.P.J Bevers, A. Prunetto, A.M. Anduaga, B. Hollis, V. Gardeux, V.S. Braman, J.M.C. Russeil, S. Kadener, M. Dal Peraro, **B. Deplancke**. Extensive tissue-specific expression variation and novel regulators underlying circadian behavior, ***Science Advances, 7***:eabc3781, 2021.

R.P.J. Bevers*, M. Litovchenko*, A. Kapopoulou, V.S. Braman, V.M. Lemos Da Silva, M.R. Robinson, J. Auwerx, B. Hollis, **B. Deplancke**. Mitochondrial haplotypes affect metabolic phenotypes in the *Drosophila* Genetic Reference Panel, *bioRxiv,* 2018; ***Nature Metabolism, 1***:226–1242, 2019.

W. Chen*, P.C. Schwalie*, E.V. Pankevich, C. Gubelmann, S.K. Raghav, R. Dainese, M. Cassano, M. Imbeault, S. Min Jang, J. Russeil, T. Delessa, D. Trono, C. Wolfrum, **B. Deplancke**. ZFP30 promotes adipogenesis through the KAP1-mediated activation of a retrotransposon-derived *Pparg2* enhancer, ***Nature Communications, 10***:1809, 2019.

P.C. Schwalie*, H. Dong*, M. Zachara*, J. Russeil, D. Alpern, N. Akchiche, C. Caprara, W. Sun, K.U. Schlaudraff, G. Soldati, C. Wolfrum#, **B. Deplancke#**. A stromal cell population that inhibits adipogenesis in mammalian fat depots, ***Nature***, 559:103–108, 2018.

Isakova, R. Groux, M. Imbeault, P. Rainer, D. Alpern, R. Dainese, A. Giovanna, D. Trono, P. Bucher, **B. Deplancke**. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. ***Nature Methods, 14***:316–322, 2017

B. Deplancke#, D. Alpern, V. Gardeux. The genetics of transcription factor DNA binding variation. ***Cell, 166***:538–554, 2016.

S.M. Waszak*, O. Delaneau*, A.R. Gschwind, H. Kilpinen, S.K. Raghav, R.M. Witwicki, A. Orioli, M. Wiederkehr, N.I. Panousis, A. Yurovsky, L. Romano-Palumbo, A. Planchon, D. Bielser, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, N. Hernandez, A. Reymond, **B. Deplancke#**, E.T. Dermitzakis#. Population variation and genetic control of modular chromatin architecture in humans, ***Cell, 162***:1039–1050, 2015

H. Kilpinen*, S.M. Waszak*, A. Gschwind*, S.K. Raghav, R.M. Witwicki, A. Orioli, E. Migliavacca, M. Wiederkehr, M. Gutierrez-Arcelus, N. Panousis, A. Yurovsky, T. Tappalainen, L. Romano-Palumbo, A. Planchon, D. Bielser, J. Bryois, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, L.J. Core, J.T. Lis, N. Hernandez#, A. Reymond#, **B. Deplancke#**,

E.T. Dermitzakis#. Coordinated effects of sequence variation on chromatin structure, DNA binding and transcription, ***Science, 342***:744–747, 2013.

J. Simicevic*, A.W. Schmid*, P. Gilardoni*, B. Zoller, S.K. Raghav, I. Krier, C. Gubelmann, F. Lisacek, F. Naef, M. Moniatte#, **B. Deplancke#**. Absolute quantification of transcription factors during cellular differentiation using multiplexed, targeted proteomics, ***Nature Methods, 10***:570–576, 2013

S.K. Raghav*, S.M. Waszak*, I. Krier, A. Isakova, C. Gubelmann, T.S. Mikkelsen, and **B. Deplancke**. Integrative genomics identifies SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPb and KAISO, ***Molecular Cell, 46***:335–350, 2012

A TECHNOLOGY-CENTRIC VIEW OF HOW OUR GENOME ENCODE CELLULAR AND PHENOTYPIC DIVERSITY

B. Deplancke[1,2]

Summary

Our genome is a sublime, but also mysterious biological entity that, despite being shared among all cells of our body, gives rise to enormous cellular heterogeneity. Moreover, the genome of one human is on average 99.9% identical to that of an unrelated individual, yet there is great phenotypic variation in the human population, for example in height, fitness, aging and susceptibility to disease. The overarching, underlying question is how does our genome manage to produce such stunning cellular and phenotypic diversity, even though the set of genetic instructions that is encoded by our genome is virtually identical between cell types and mostly so between human individuals? This is clearly a fundamental question in biology and one that has already occupied numerous labs in the world, including my own, for many years. While initially hampered by a general lack of adequate experimental and analytical tools, efforts to better understand the function of our genome were greatly boosted by the availability of the human genome DNA sequence at the turn of this century. This in turn spurred a technological and analytical revolution, resulting in the development of tools and resources that allowed us to holistically probe our genome in the context of development, homeostasis and disease, first at the bulk tissue level, but increasingly in recent years at the single cell level. In this review, I summarize how in the now 15 years of its existence, my lab contributed to this revolution, developing both novel methods and analytical tools that allowed us to gain new insights into the inner workings of this beautiful set of DNA molecules: our genome.

---

1 Institute of Bio-engineering, School of Life Sciences, EPFL, Station 19, CH-1015 Lausanne

2 Swiss Institute of Bioinformatics, CH-1015 Lausanne

Introduction

The genome, our book of life

Our genome is a most beautiful book, one that: "wrote itself, continually adding, deleting and amending over four billion years" (Matt Ridley, Author of "The Genome")

The genome is the fascinating genetic material that one can find in almost all cells of the body. We know that it contains all the instructions that are required to generate and maintain a specific organism, which is why it is often referred to as the book, or even better yet, the manual of life. Yet, how these instructions are encoded in the genome, also known as "the regulatory code", and how they are interpreted and executed by the non-DNA machinery in a cell remains poorly understood. To crack this regulatory code, a first requirement is of course to be able to assess the content of the manual. This is why the release of the nearly complete genome sequences of humans and model organisms such as the mouse and *Drosophila* in the beginning of the 21st century was such a fundamental turning point as it for the first time provided a glimpse at all the parts that may be required to build a complex organism such as a human, including the protein-coding genes, but also the so-called non-coding sequences that may control which genes are expressed where, when and at which level. Considerable progress has since been made to annotate the ~25,000 genes encoded by both the human and mouse genomes. However, much less is still known about the sequences, i.e. regulatory elements, and corresponding networks that control the expression of these genes. This is because we now realize that the regulatory code is cryptic and degenerate, in contrast to the genetic code, which renders it very difficult to infer function from sequence information alone.

In my lab, we have had a continuous interest in dissecting how readers of this regulatory code – proteins that are called transcription factors (TFs) and their co-regulators – work together and interact with the genome to interpret its embedded set of instructions that allow a cell to change its identity and for example become a new cell type (Chen et al., 2019; Deplancke, 2009; Gubelmann et al., 2014; Pradhan et al., 2017a; Raghav et

al., 2012). This is conceptually comparable to seeing a beautiful figure made out of distinct Lego blocks (the "genes") and trying to decipher what the instructions could be that enabled a person to create this figure, i.e. putting the appropriate type of Lego blocks on top of one another in the right configuration and right color (or "expressing the genes at the right time, level, and place") (**Fig. 1**). To assist the Lego enthusiast, a manual is typically provided in the Lego box, guiding the user to build his/her creation in step-wise fashion. Of course, our genome does not come with a manual and it is up to us, scientists, to reverse engineer its instructions. In other words, given a specific cell type, can we infer the underlying manual, i.e. its core parts, starting with regulatory elements and TFs, and how these parts together give rise to that cell type. The interactions between regulatory elements and TFs constitute gene regulatory networks (GRNs). The ultimate function of these networks is to coordinate the progression of distinct regulatory states in space and time, which lies at the heart of nearly all biological processes such as differentiation, cycling, responses to environmental stimuli etc. Thus, to understand how GRNs orchestrate differential gene expression programs underlying a biological process of interest and essentially define the manual of life, it is crucial to devise technologies that allow us to identify all implicated regulatory elements and corresponding TFs as well as how, when, and where they interact, a great challenge to which our lab has consistently contributed contributed (Alpern et al., 2019; Gubelmann et al., 2013; Hens et al., 2011; Isakova et al., 2017; Simicevic et al., 2013)(Dainese et al., 2020).

Fig. 1. The Lego Analogy. *While the fundamental building blocks (Lego pieces) are the same between these distinct creations, by tinkering with the number, shape and color of these pieces and with when they are used during the building process, one can create vastly different Lego animals. Similarly, while the genes are the same in each cell, by modulating which genes are expressed, where and when, very different cell types can be generated. This analogy even holds for within and between species' comparisons. Indeed, even between human and mouse, > 80% of the genes are shared, which is why already back in 1975, King and Wilson, when comparing human and chimpanzee genes, argued in their seminal Science paper that "their macromolecules are so alike that regulatory mutations may account for their biological differences" (King and Wilson, 1975). In other words, it is the regulatory code, the manual of life, that drives cellular heterogeneity and variations in this code result in phenotypic diversity and can even account for species' differences. Copyright@ Legotruman.*

notypic variation. Imagine that someone makes slight modifications in the Lego manual: how would the resulting Lego figure look like? The same? Better or worse? It turns out that in nature, all three scenarios are possible, but how a slight modification in the manual can induce an altered phenotype, e.g. differential disease susceptibility, is again poorly understood. This is not surprising given that, as indicated above, we still have a hard time deciphering the manual of life for one cell type, let alone the development of an entire organism or a variation thereof! Addressing this fundamental question is further complicated by the fact that environmental factors may also play a role in affecting a specific phenotype (e.g. smoking increases your risk of lung cancer). This is why in my lab, we

also study the effect of genomic variation not only in humans (Kilpinen et al., 2013; Waszak et al., 2015), but also in model organism populations such as that of the fly (Bevers et al., 2019; Bou Sleiman et al., 2020, 2015; Frochaux et al., 2020; Litovchenko et al., 2021) given that we can raise these animals in highly standardized settings, effectively eliminating environmental confounders.

Adipose tissue: our blanket of life

"The devil has put a penalty on all things we enjoy in life, either we suffer in health, or we suffer in soul, or we get fat" (Albert Einstein)

Humans, as other mammals, benefit from a subcutaneous fat layer that insulates and cushions the outer world, hence the long-lasting perception that this layer serves merely as a comfort blanket. However, the rapid rise in obesity, reaching pandemic levels, and especially its many co-morbidities such as diabetes, cardiovascular disease and even cancer, have forced the scientific community to revisit the physiological importance of adipose tissue. This has led to a true scientific (re)valorization of this prevalent tissue, recognizing not only its biomechanical and insulating roles but also its key involvement in systems metabolism & physiology, tissue growth, regeneration and repair, as well as innate immunity.

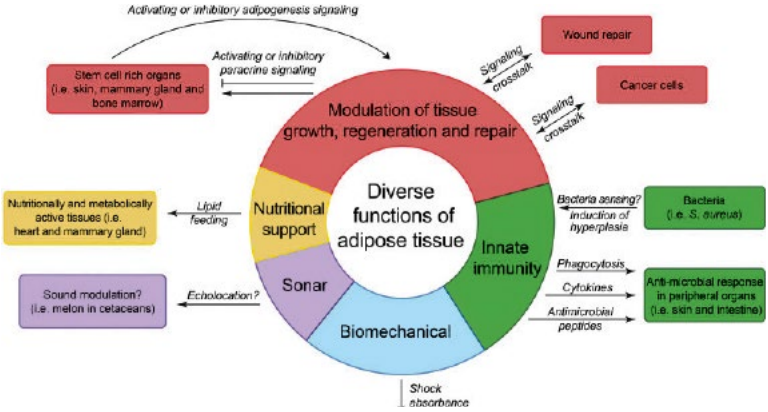

Fig. 2. Key functions of adipose tissue. *From (Zwick et al., 2018).*

While all of us can easily locate our respective fat layers and clearly recognize the biomedical importance of studying adipose biology, it may come as a surprise that we still have a relatively poor understanding of how adipocytes are formed, both in terms of their respective stem / precursor origin and the structure and function of adipogenic regulatory networks. This is partially due to the highly heterogeneous and unstructured nature of adipose tissue depots, which are present in multiple anatomical locations and consist of a mixture of different cell types, whose origin and identity differs between distinct fat depots. Nevertheless, adipogenesis is still one of the best studied differentiation paradigms, greatly aided by suitable *in vitro* models such as the mouse 3T3-L1 cells that closely mimic the molecular processes underlying especially the terminal phase of adipocyte differentiation. Consequently, much progress has been made in at least defining a core GRN that controls 3T3-L1 terminal adipogenesis. The first step in this process is triggered by the TFs C/EBPβ and δ, which are induced immediately after the addition of pro-adipogenic stimuli, but are initially inactive. Once a subsequent C/EBPβ-dependent clonal cell expansion round is completed, then these TFs activate the transcription of the pro-adipogenic master regulators PPARγ and C/EBPα. Indeed, the latter TFs are thought to control the expression of adipogenic genes and genes mediating cell cycle withdrawal. In addition, they also cross-regulate each other effectively maintaining a terminal differentiation state. Finally, no other factor has thus far been identified that can promote adipogenesis in the absence of PPARγ. However, while PPARγ is necessary, it is clearly not sufficient, implying the involvement of several other TFs that need to be integrated in the overall adipogenic GRN. This constituted the state of the art when back in 2007, we took it upon us to devise new technologies that would improve our ability to map mammalian GRNs, aiming to then specifically apply these technologies to study the GRNs' underlying adipogenesis. Little did we know then that this quest would take us well beyond the mapping of GRNs alone. Indeed, driven by new revolutionary single cell omic methods, we ventured into resolving adipose stem cell and precursor heterogeneity, uncovering a new cell type that may well fundamentally change our view of how adipose tissue development and homeostasis is controlled, as will be detailed below.

Thus, the principal questions that my lab aims to address are:

1. How the genome steers the development of specific cell types and adipocytes in particular.
2. How variation in this genome makes each of us different.

With this manuscript, I intend to review the technological contributions that my lab has made to addressing these questions as well as the scientific findings that emerged from our studies with respect to understanding genome function. I thereby restricted the biological scope of this review to our studies involving mouse and human models with a specific focus on our adipose biology work, which was honored with the Cloëtta Prize. I intentionally restricted the listed references primarily to our own papers, referring the reader for a more comprehensive and impartial representation of the pertinent literature to already published reviews of my group (Deplancke, 2009; Deplancke et al., 2016; Ferrero et al., 2020; Pradhan et al., 2017b) as well as references in our publications that are listed here.

Development of technologies to map (adipogenic) gene regulatory networks

The state of a cell is defined by its gene and ultimately protein expression profile, which itself depends on how the cell's entire regulatory circuit or network is wired at that point in time. Transitions to a different cell state (e.g. during differentiation or re-programming) are then mediated by transcriptional changes through circuit re-wiring. Thus, to understand cell behavior and function, we need to achieve a qualitative and quantitative understanding of the structural and dynamic properties of the underlying gene regulatory networks (GRNs) by identifying all implicated nodes (TFs and regulatory elements), and map the dynamic interactions, i.e. the regulatory edges, between them. To do so, we followed two complementary strategies:

i) Gene-centered:

The first strategy is gene-centered and aims to screen adipogenic regulatory elements (e.g. promoters, enhancers) and especially those linked to TF-coding genes for interacting TFs using a high-throughput version of

the yeast one-hybrid (Y1H) assay that I developed during my post-doctoral studies (Deplancke et al., 2004; Deplancke et al., 2006). The Y1H system is conceptually similar to the better-known yeast two-hybrid system, except that a DNA fragment is used as a bait together with a single hybrid protein. The DNA bait is cloned upstream of a Y1H reporter gene (e.g. *HIS3*) and integrated into the yeast genome. Thus, regulatory elements can be tested in a "more innate" chromatinized format for their ability to associate with specific TFs. This format increases the specificity, reduces the number of false positives, and therefore provides a straightforward validation method for *in vitro* or *in silico*-derived protein-DNA interaction datasets (Deplancke et al., 2004). For Y1H assays to be most effective, regulatory elements are ideally directly screened versus a library of TFs, which requires the availability of a comprehensive TF ORF (open-reading frame) resource for your model system of interest. A first, major project in my lab was therefore the generation of a versatile mouse TF resource, which now already contains over 1,000 (out of ~1550, [Gubelmann et al., 2013]) fully sequence-verified TF ORF clones. To subsequently enable the yeast-based screening of elements of interest versus this array of mouse TFs, we developed a cross-platform pipeline to experimentally analyze these elements for interacting TFs at unprecedented throughput and resolution (**Fig. 3**). Key here is the implementation of a microfluidic approach, MARE for Microfluidics-based Analysis of Regulatory Elements, following the Y1H screen that enables the fine-grained localization of TFs of interest within specific regulatory elements (Gubelmann et al., 2013; Hens et al., 2011). The MARE technique can be compared with a series of electrophoretic mobility shift assays (EMSAs), in which a TF is tested for its ability to bind to a collection of typically small DNA sequences, and relative DNA occupancy data for each sequence can be derived. Similar to EMSA, the MARE protocol starts with small DNA elements, resulting from the fragmentation of long regulatory DNA sequences, which are tested individually for binding to a specific set of TFs (e.g., those that were identified using our Y1H screens). However, MARE accommodates >700 EMSA-like assays at once on one microfluidic chip in a relatively straightforward and cost-effective manner. This in turn enables the generation of a relative DNA occupancy landscape for each TF of interest over the length of the respective regulatory element where the regions of highest occupancy likely

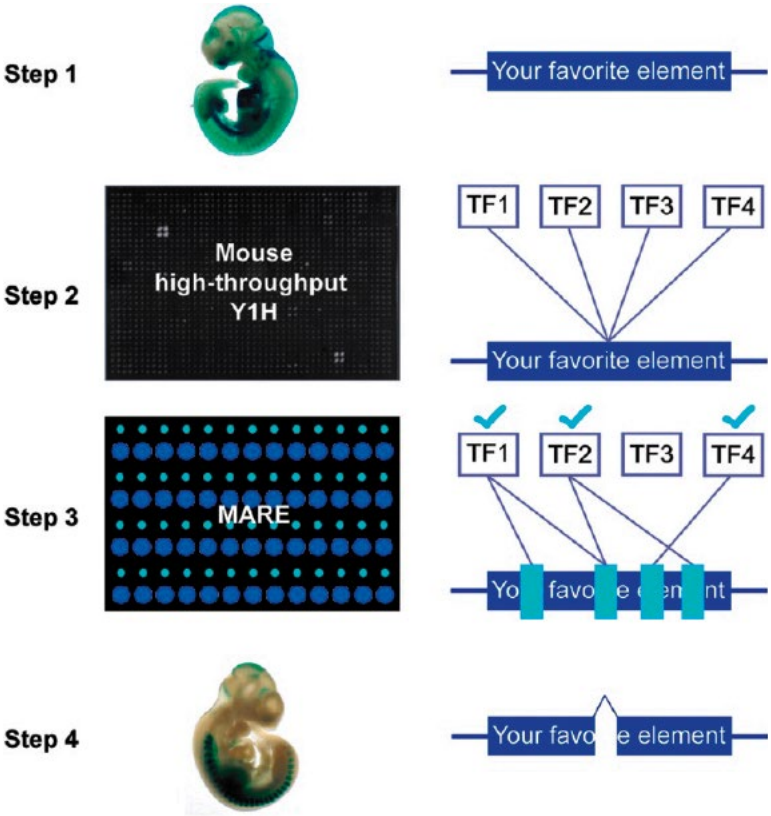

Fig. 3. Schematic overview of the pipeline employed to de-orphanize mammalian gene regulatory elements. *The regulatory element of interest is first cloned (Step 1), and then integrated into yeast to enable high-throughput Y1H screens leading to the identification of putatively interacting TFs (Step 2)(shown is a typical Y1H screening plate featuring two strongly "positive" TFs, which emerge as quadrants given that each TF is independently tested 4 times). In Step 3, MITOMI-based Analysis of Regulatory Elements (MARE, see* ***Research Aim 2*** *for more details) is performed to both validate (reflected by light blue check marks) and map the detected TF-DNA interactions (indicated by light blue boxes) within the respective regulatory element. Finally, small binding regions of interest can be deleted to examine the relevance of these DNA segments in mediating the in vivo activity of the regulatory element (Step 4). Taken from (Gubelmann et al., 2013).*

contain the respective TF binding sites. Thus, MARE allows the simultaneous validation and localization of protein-DNA interactions within regulatory elements, greatly increasing the DNA binding site resolution that Y1H screens can typically provide. As a proof-of-concept, we validated this pipeline using well-described regulatory elements and orphan enhancers, demonstrating that it enables the identification of known and novel mouse TF–DNA interactions that are relevant *in vivo* (Gubelmann et al., 2013).

ii) TF-centered:

SMiLE-seq

Our Y1H-based, gene-centered approach uniquely allows the screening of any regulatory element of interest for interacting TFs in high-throughput fashion. A clear drawback of this approach though is that, while interactions are tested *in cello* (i.e. yeast cells) in a chromatinized context, they are obviously not probed in their natural, cellular setting. In addition, having screened hundreds of regulatory elements and identified many more interacting TFs, we realized that many of these interactions could in fact be simply predicted *in silico* by computationally mapping the *in vitro-derived* DNA binding motifs of TFs to target sequences of interest. In other words, by analyzing regulatory elements of interest for the occurrence of small, often 8–12 bp long motifs that correspond to DNA binding preferences of TFs, we were able to predict the majority of TF-DNA interactions that our Y1H screen managed to pick up. Such computational analysis also suffers from disadvantages of course. For example, it is prone to false positives as such motifs are short and degenerate and it is therefore prudent to implement rather stringent motif matching parameters to avoid calling too many interactions. In addition, of the >1500 TFs that are encoded by the human or mouse genomes (making it the largest protein family), still a surprisingly high number of TFs (estimated to be >400) are still "orphan". This means that no DNA binding specificities have so far been derived for such TFs, despite the importance of characterizing these specificities to increase our understanding of the regulatory logic of a cell. The latter also rationalizes why tremendous efforts have already been invested involving various *in vitro* and *in vivo* techniques such as protein binding microarray (PBM), HT-SELEX, bacterial one-hybrid, or chromatin immunoprecipita-

tion (ChIP) coupled to sequencing (ChIP-seq) to map TF binding motifs. But despite these efforts, the current catalogue of individual, characterized TFs, let alone TF heterodimers or higher complexes, is still vastly incomplete and lacks quantitative insights. This realization led us to explore alternative protein-DNA interaction profiling strategies that would not only prove highly robust, quantitative, and superior to other, comparable assays in key DNA binding characterization parameters, but also semi-automated, fast, and thus easily implementable.

Leveraging our expertise in microfluidics and building on a clever experimental platform, MITOMI for Mechanically Induced Trapping Of Molecular Interactions that was developed by my colleague here at the EPFL, Prof. Sebastian Maerkl, while at Prof. Steve Quake's lab in Stanford, we developed a new TF DNA binding characterization technology, SMiLE-seq for Selective Microfluidic Ligand Enrichment coupled to sequencing (Isakova et al., 2017) (**Fig. 4**). We specifically engineered SMiLE-seq to complement existing TF-DNA binding profiling approaches by 1) producing binding models that tend to have superior predictive power compared to other *in vitro* models; 2) enabling the characterization of the DNA binding preferences of both monomers, homo- and heterodimers, and 3) studying the DNA binding specificities of TFs (e.g. C2H2-containing zinc fingers) that have so far consistently resisted comprehensive DNA binding characterization by other *in vitro* methods.

Given its ability to de-orphanize, difficult to probe TFs, SMiLE-seq is now being enlisted to investigate the remaining, roughly 400 uncharacterized human TFs, aiming to complete at least the "TF code". These efforts are currently part of a large consortium project led by Prof. Tim Hughes, called "Codebook", in which these 400 TFs are subjected to most available protein-DNA interaction profiling methods including protein-binding microarrays, HT-SELEX, ChIP-seq and thus SMiLE-seq. The resulting data resource, to my knowledge the largest of its kind to date, should be of great value, not only to bring the human TF motif catalogue near completion, but to also directly compare the strengths and weaknesses of the different methods themselves.

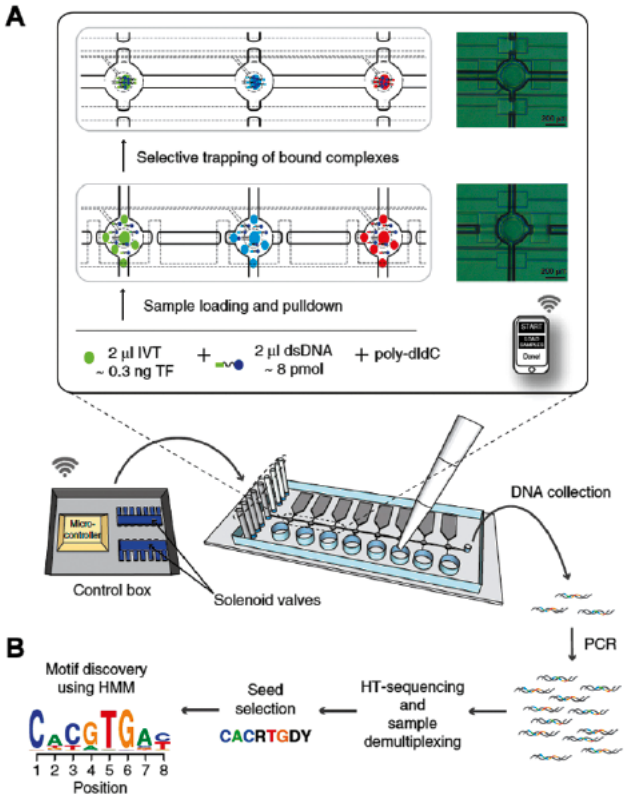

Fig. 4. *Adapted from (Isakova et al., 2017). A. Schematic representation of the experimental SMiLE-seq setup.* ***A*** *snapshot of three units of the microfluidic device is shown. In vitro transcribed and translated (IVT) bait TF, target dsDNA, and a nonspecific competitor poly-dIdC are mixed and pipetted in one of the wells of the microfluidic device. The mixtures are then passively pumped in the device for 20 min (bottom panel). Newly formed TF–DNA complexes are trapped under a flexible polydimethylsiloxane (PDMS) membrane, and unbound molecules as well as molecular complexes are washed away (upper panel). Left, schematic representation of three individual chambers. Right, corresponding snapshots of an individual chamber taken before and after mechanical trapping.* ***B.*** *TF motif discovery pipeline. The bound DNA is eluted from all the units of the device simultaneously and collected in one tube. Recovered DNA is amplified and sequenced as a 2–4% spike-in. The sequencing reads are then demultiplexed, and the seed sequence is identified for each sample (here using the algorithm MEME). This seed is then used as an input reference sequence for Hidden Markov Modeling (HMM)-based TF motif discovery.*

While the limitation of the Y1H assay is that protein-DNA interactions are probed outside their natural context, that of motif-based approaches is that any prediction is purely based on an *in silico* analysis. This is why a technology such as ChIP-seq has always remained the most popular protein-DNA interaction detection tool since it probes such interactions *in vivo* and provides thus direct read-outs of which TFs might be targeting which genes. Clear disadvantages of ChIP-seq, rationalizing the existence of other protein-DNA interaction detection tools, is the limited resolution as to where exactly a TF might be binding to DNA (although a rather novel tool such as exo-ChIP-seq managed to vastly improve this), the inability to demonstrate that the detected interaction is direct (as it could be mediated by collaborating TFs), and the slow and laborious nature of the experimental process. In my lab, we were routinely performing ChIP assays to profile the chromatin landscape in various cell types with the goal of identifying and characterizing regulatory elements or domains that may control gene expression in our biological processes of interest. Over time, we became however increasingly frustrated with the limited throughput and sensitivity of these ChIP assays as well as their high complexity. This prompted us to look for alternate solutions that would allow us to substantially simplify these ChIP-seq procedures. Building on our expertise in developing SMiLE-seq (Isakova et al., 2017), we started to investigate whether the SMiLE-seq approach used to trap and isolate DNA bound to *in vitro* expressed TFs could also be used to profile DNA bound by *in vivo* chromatin marks or TFs. Excitingly, after extensive microfluidic chip redesign and some additional tweaks and turns, we managed to develop a microfluidic ChIP-seq platform, dubbed "FloChIP" (Dainese et al., 2020), which can now be used to perform high quality epigenomic or TF binding assays on samples as small as 100 cells. In addition, the workflow is 10-20-fold faster than other available solutions, also because of its inherent modularity, for example by enabling to probe up to eight chromatin marks in parallel, starting from the same sample **(Fig. 5A)** or 8 samples in parallel, targeting one chromatin mark **(Fig. 5B)**. Finally, the platform also allows to sequentially ChIP samples, as such enabling for example to study the importance of bivalent marks in gene regulation.

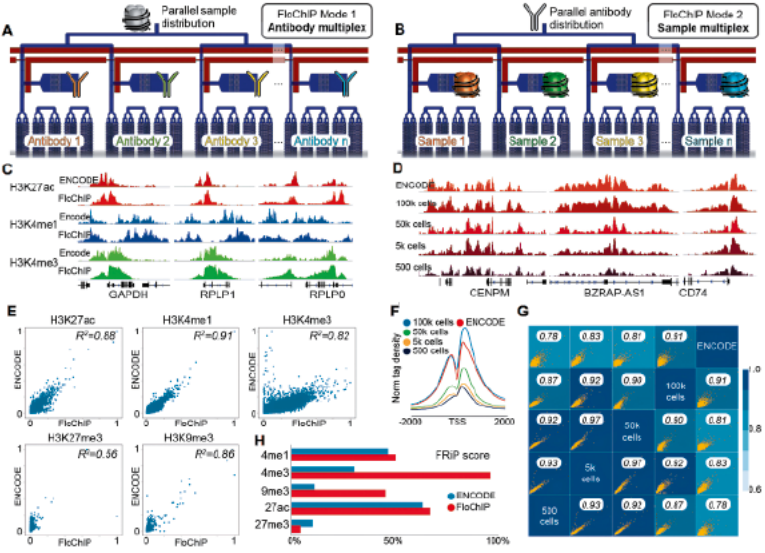

Fig. 5. A. *Schematic depiction of FloChIP's mode 1: antibody multiplex. Each IP lane is functionalized separately by introducing different antibodies through the individual inlets. During IP, one sample is introduced through the common inlet and distributed equally across all IP lanes, hence enabling multiple IPs involving distinct antibodies at once. B. Schematic depiction of FloChIP's mode 2: sample multiplex. One antibody solution is introduced through the common inlet and distributed equally across all IP lanes. During IP, each IP lane is loaded separately by introducing different samples through the individual inlets. Taken from (Dainese et al., 2020).*

Targeted, quantitative proteomic analysis of (adipogenic) TFs – TF DNA binding modeling

ChIP-seq provides valuable insights into the DNA binding landscape of focal TFs, yet how such landscapes are established remains relatively poorly understood, but are thought to at least partially reflect the DNA binding affinity and nuclear concentration of TFs. Consequently, deriving nuclear TF copy numbers has been of longstanding interest in regulatory genomics, but only few studies have so far provided estimates on the absolute *in vivo* abundance of TFs. This can be explained by the fact that TFs tend to be lowly expressed (at least compared to many other protein families), rendering them difficult to identify and specially to quan-

tify with standard shotgun liquid chromatography-mass spectrometry (LC-MS) approaches.

To alleviate this issue, we turned to a particularly sensitive MS-based technique termed Selected Reaction Monitoring (SRM), which features an excellent sensitivity by only targeting a subset of detectable peptides that are specific to the protein of interest (i.e. proteotypic peptides). Thus, for SRM to work optimally, proteotypic peptides for target proteins need to first be identified or derived. The selection of proteotypic peptides for TFs proved however challenging, mainly because of the scarcity of TF peptide data in public repositories and the difficulty of detecting TF-specific tryptic peptides in discovery experiments. We therefore decided to adopt an *in vitro* full-length protein expression-based strategy to initiate the construction of a mammalian TF-specific proteotypic peptide atlas (Simicevic et al., 2015), capitalizing on the availability of a comprehensive TF ORF clone library in the lab, as described above (Gubelmann et al., 2013). Building on these efforts, we then developed a state-of-the-art, targeted SRM-based assay, which combines high sensitivity and technological innovation to enable the monitoring of absolute copy number changes of TFs of interest during specific biological processes using *in vitro*-expressed, isotopically-labeled protein standards (**Fig. 6**).

As a proof-of-concept, we quantified in absolute amounts the levels of the adipogenesis master regulators PPARγ and RXRα in the nucleus of 3T3-L1 cells at six time-points during adipogenesis, and subsequently inferred their nuclear copy number per cell. We then used these data to build a quantitative model of genome-wide TF DNA binding in collaboration with our EPFL colleague Prof. Felix Naef. The goal here was to extend already available GRN models by incorporating our rather unique TF abundance data, allowing for better model calibration and thus prediction of dynamic network changes. Next to our microfluidics (MITOMI)-based biophysical characterization of PPARg's protein and DNA binding properties (Isakova et al., 2016) the ensuing model provided unique, quantitative insights into *in vivo* PPARg DNA binding. Specifically, we revealed that PPARγ's DNA binding profile can be faithfully modeled by considering its own copy number, thermodynamic principles, and chromatin accessibility. The functional consequence of our findings is that the chromatin state appears to constitute a "landing map" for

PPARg DNA binding, thus emphasizing the importance of both protein copy number and chromatin remodeling in dictating TF DNA binding behavior during adipogenesis (Simicevic et al., 2013).

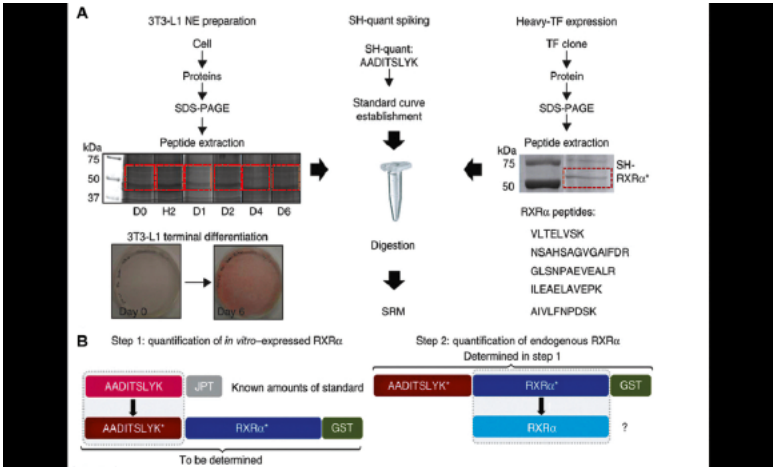

Fig. 6. A. *Left, preparation of 3T3-L1 total nuclear protein extract (NE). Cells are lysed at the indicated differentiation time point (D0–D6; D, day; H, hour), after which nuclear proteins are extracted. The resulting protein mixture is separated by SDS-PAGE, and TF bands are excised from the gel. Right, preparation of in vitro-expressed SH-tagged TFs. The constructs are expressed as heavy-labeled versions (*), purified by glutathione S-transferase (GST) affinity and separated by SDS-PAGE. Bands containing the heavy-labeled constructs, here SH-RXRα-GST*, are excised from the gel. Center, each nuclear extract band to be quantified is mixed with a gel slice of the in vitro-expressed TF construct, spiked with known amounts of light SH-quant tag and digested in gel. SH-quant features a C-terminal trypsin-cleavable fluorescent tag (here termed JPT) that is used to quantify this quantotypic peptide. The resulting peptide mixtures are quantified by SRM using proteotypic peptides selected by performing shotgun mass spectrometry analyses on each in vitro-expressed TF. Quantification of each TF requires a separate experiment in this configuration.* ***B.*** *Schematic of the quantification approach as outlined in* ***A.*** *Taken from (Simicevic et al., 2013).*

TF overexpression screen

The TF-centric approaches introduced above enable the characterization or quantification of TFs of interest, yet are unable to provide comprehensive insights into the TFs and underlying GRNs that mediate a specific biological process such as adipogenesis. To address this, we devised a high-throughput screening strategy involving both our mouse TF ORF collection which was transferred to a lentiviral overexpression vector and a robotic screening and imaging platform to systematically evaluate the functional involvement of TFs in fat cell differentiation. These experiments, performed in collaboration with the labs of Profs. Didier Trono (EPFL) and Christian Wolfrum (ETHZ), revealed 26 (3.5%) and 39 (5%) TFs that significantly enhance or repress adipogenesis, respectively, constituting a rich catalog to further dissect the adipogenic GRN. Intriguingly, while PPARγ was identified within the top 10 pro-adipogenic TFs, validating our screen, very little was known about the top three enhancing TFs: ZEB1, ZFP30, and ZFP277.

Follow-up molecular validation experiments revealed that the latter top pro-adipogenic TFs are indeed required in adipogenesis. Specifically, for ZEB1, we found that this well-known EMT TF directly targets virtually every TF that has so far been mapped within the adipogenic GRN and the majority of these TF-coding genes decrease in expression upon ZEB1 knockdown ([Gubelmann et al., 2014]; **Fig. 7**). This identifies ZEB1 as a new adipogenic master regulator with likely involvement in both the commitment and terminal differentiation phases given its high expression at these time points. Consistent with the latter hypothesis is the observation that knockdown of ZEB1 significantly reduces the expression of commitment markers and also inhibits mesenchymal stem cell differentiation to adipocytes. We further found evidence that ZEB1 may mostly act as a co-activator of gene expression through its association with the early adipogenic TF C/EBPb, which we had previously revealed primes adipogenic regulatory elements for activation through its association with the co-repressor and gatekeeper Nuclear Receptor Co-Repressor 2 (Raghav et al., 2012).

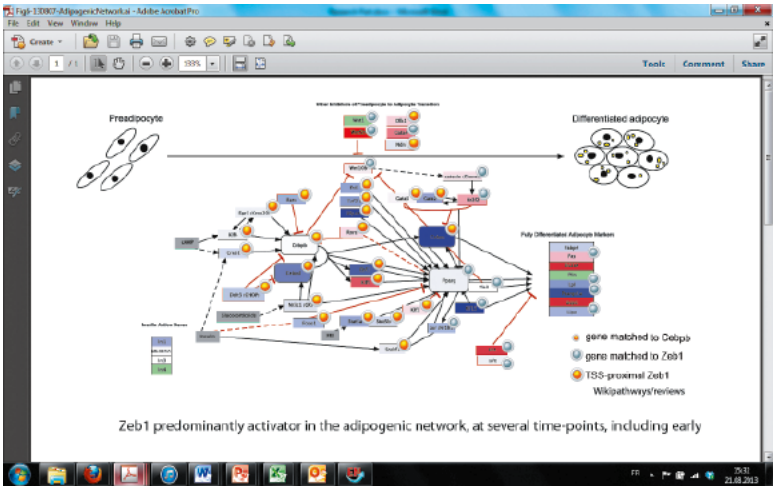

Fig. 7. I*ntegration of ZEB1 within the established core adipogenic GRN through RNA-seq and ChIP-seq-based analyses. Genes that are up – or downregulated upon ZEB1 knock-down are highlighted in respectively red and blue (with the extent of change scaling with color intensity). The data shown are retrieved at D0 of differentiation.*

For ZFP30, we had to start from scratch since virtually nothing was known at the time about the function of this TF. Domain analysis revealed however that ZFP30 belongs to the family of KRAB domain containing zinc finger proteins (KZFPs). Despite their abundance, most of the KZFPs have not been functionally characterized, although they are generally thought to be involved in the repression of transposable elements (TEs) through their association with the co-repressor KAP1. Thus, by studying the transcriptional function of ZFP30, we could contribute both to dissecting the adipogenic regulatory network and to elucidating the role of co-repressors within this system. Using *in vitro* and *in vivo* assays in mouse and human, we were able to define the mechanism of ZFP30-mediated adipogenic activation**,** revealing that the expression of the master adipogenic regulator *Pparg2* is controlled by an ancient retrotransposon-derived enhancer targeted by ZFP30 (**Fig. 8**). This potent regulatory sequence is over 100 million years old and derived from an L1 retrotransposon that was likely under repressive control of ZFP30 in the ancient past. ZFP30 was thus co-opted into the local eutherian adipogenic regu-

latory network, influencing the expression of its very master regulator, *Pparg*2. Intriguingly, however, unlike the canonical role of KZFPs as repressors of retrotransposon elements, we found that ZFP30 activates the *Pparg2* enhancer by maintaining its KRAB-mediated interaction with KAP1, but Ser-473 phosphorylation of KAP1 appears to mediate a switch from the canonical co-repressor to a co-activator function. Consistently, we demonstrate that loss of KAP1/phosphorylation negatively affects adipogenesis, revealing a context-specific regulatory function for KAP1 in driving fat cell differentiation.

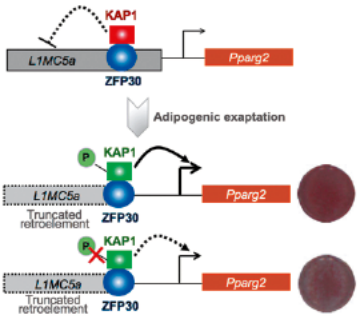

Fig. 8. *Graphical abstract illustrating the proposed molecular mechanism underlying the pro-adipogenic function of ZFP30. The latter TF was likely originally recruited to the Pparg locus to repress the activity of the L1MC5a retroelement. The subsequent insertion of a murine-specific retroelement (B3A, not shown) truncated the first retroelement, rendering the repressive function of ZFP30 obsolete. Subsequently, we hypothesize that ZFP30 was adopted into the local adipogenic regulatory network, as mediated by the phosphorylation of its co-regulator KAP1 which turned this complex into an activating rather than repressing entity. Consequently, loss of KAP1 phosphorylation abrogates its activating capacity, resulting in reduced Pparg2 expression and less adipogenesis (as visualized by Oil-red-O staining of differentiation-induced 3T3-L1 cells).*

Leveraging the resolving power of single cell transcriptomics to study adipogenesis

Our work on ZEB1 revealed that, while great advances were made to understand the terminal phase of adipogenesis, much less was known about the molecular drivers underlying adipogenic commitment. This in part reflected the lack of universally accepted markers that are either specific for one adipocyte precursor type or uniformly label a particular precursor population, rendering the identity of adipocyte stem and precursor cells (ASPCs) still enigmatic. We realized that this is exactly the type of problem that could be very effectively addressed by single cell transcriptomics (scRNA-seq) given that it has enormous potential to contribute to

our understanding of cell type diversity, tissue structure and homeostasis, development, and pathology. In collaboration with our collaborator Prof. Wolfrum (ETHZ), we therefore set out to acquire for the first time an scRNA-seq-based, high-resolution view of cellular heterogeneity within the stromal vascular fraction (SVF) of fat depots. To do so, we profiled and characterized FACS-isolated CD34+, CD29+, SCA1+, Lin- C56BL/6J mouse subcutaneous fat SVF cells, which are widely regarded as most closely resembling ASPCs. Our initial analysis involved 191 mouse ASPCs from three independent experiments using the Fluidigm C1 system, revealing ~5,000 genes that were expressed on average in each cell ([Schwalie et al., 2018]; **Fig. 9A**). To examine whether these ASPCs constituted one large or several subpopulations, we performed *de novo*, unsupervised clustering based on cell-derived gene expression profiles. Interestingly, this analysis revealed three main subpopulations. These were not only largely validated using another independent analysis involving now more than 1,804 Lin-cells that were processed using the 10X Genomics Chromium dropleting system, but they were also discovered in subsequent analyses by independent research groups, as shown through a meta-analysis by our group in (Ferrero et al., 2020).

The identification of these scRNA-seq-based subpopulations suggested functional differences among ASPCs, prompting us to phenotype the different cell types *in vitro* and *in vivo*. We analyzed several surface markers that are (relatively) specific to one of the three subpopulations and that thus could allow subpopulation-specific cell isolation and characterization. Using the marker CD142, encoded by the gene *F3*, we were able to highly enrich for the blue population (**Fig. 9A**). Intriguingly, and contrary to their presumed multipotent nature, subsequent differentiation analyses revealed that these CD142+ cells are largely refractory to adipogenesis (**Fig. 9B–C**). Moreover, depletion of these cells from the initial pool of ASPC-enriched SVF appeared to highly enhance adipogenesis, suggesting that CD142+ASPCs have inhibitory properties (**Fig. 9B–C**), an effect that could be replicated *in vivo* (**Fig. 9D-E**).

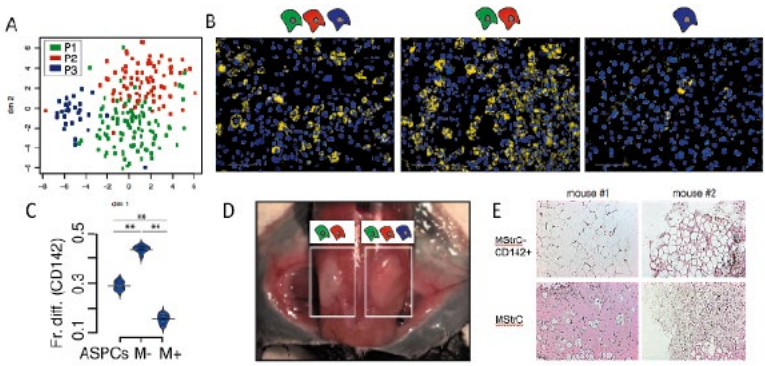

Fig. 9. A. *Unsupervised clustering of 191 ASPCs from C57BL/6J mouse subcutaneous fat SVF using tSNE revealed three major subpopulations (P) (represented as different colors) that exhibit significant differences in gene expression.* ***B.*** *Microscopic images of distinct ASPC fractions (total ASPCs, left; CD142- ASPCs, middle; CD142+ ASPCs, right) after adipogenic differentiation. Nuclei are stained with DAPI (blue) and lipids with Bodipy (yellow).* ***C.*** *Beanplots showing the distribution of the fraction of differentiated cells per each ASPC fraction (enriched or depleted in cells that feature the CD142 marker (M)) shown in* ***B.*** *(* p-value <=0.05, ** p-value <=0.01, t-test).* ***D–E.*** *Histological images of matrigel implant plugs composed of either total ASPCs or CD142-ASPCs after three weeks of high-fat diet feeding (adapted from (Schwalie et al., 2018)).*

To our knowledge, the existence of such a cell type within a specific population of cells (in this case: Lin- CD29+ CD34+ SCA1+) had so far never been reported, except for T regulatory cells (Tregs). In the immune system, Tregs use their immunosuppressive capacities to maintain immune homeostasis and mediate peripheral tolerance. We therefore proposed to name this newly uncovered cell population “Aregs” (for Adipogenesis Regulators) and provided in our study its first extensive anatomical and molecular characterization: we show that Aregs reside around the vasculature and are typified by the expression of coagulation and complement cascade factors that are recognized modulators of adipose tissue. These collective findings have great biomedical implications in metabolism and beyond because: 1) adipose tissue mass can expand both by hyperplasia (increase in cell number) and hypertrophy (increase in cell mass). Since the former is widely regarded as metabolically “healthier obesity” compared to the latter, it is of great interest to understand why obese individuals frequently experience overtime this mode shift in fat mass expan-

sion. One recurrent hypothesis is that hypertrophy may be triggered or enhanced upon depletion of the ASPC pool after chronic overfeeding. However, our findings now suggest that this mode shift might not be solely due to alterations in the number of ASPCs, but also to the amount of Areg cells, as these could control the tissue's *de novo* adipogenic capacity; 2) mesenchymal stromal cell-driven adipogenesis is routinely performed in thousands of labs around the world, yet it remains entirely unclear why overall fat cell formation is relatively inefficient, rarely reaching >50% cell differentiation. Our results now indicate that there is a thus far unrecognized cell population among these isolated cells that actively blocks this differentiation process. Whether a similar concept could be applicable to other differentiation systems is an exciting prospect; 3) it is well established that adipocytes can arise in other tissues such as bone marrow and muscle, albeit it remains unclear why these cells only form under specific (patho)physiological conditions. Based on our data, it is tempting to speculate that in these tissues, *de novo* adipogenesis is controlled not only through the presence or absence of adipocyte precursors, but also through the presence of Aregs. Thus, our findings point to a critical role for Aregs in modulating the plasticity and metabolic signature of distinct fat-cell containing systems, where they may constitute essential components of the elusive adipogenic precursor niche.

At this point, we still have a very poor understanding of the molecular nature and function of these enigmatic Lin-CD34+SCA1+CD142+ cells and whether they are functionally conserved in humans. In a recent, follow-up study (Zachara et al., 2021), we therefore aimed to provide multi-omics- and experiment-based insights into the molecular mechanisms that control the developmental emergence and function of at least mouse Aregs. As summarized in **Fig. 10**, we found that Aregs constitute a clearly distinct and stable CD142+ ASPC subpopulation in adult mice, which remains phenotypically robust regardless of experimental conditions, such as the source of anti-CD142 antibodies, various cell sorting gating strategies, the strength of adipogenic differentiation cues, or even sex of the animals. In addition, we uncovered unexpected developmental dynamics since, contrary to adult Aregs, pre-weaning CD142+ ASPCs exhibit a high adipogenic propensity. Indeed, they acquire *bona fide* Ar-

eg-like molecular and functional properties only during the third post-natal week, an event possibly triggered by weaning or sexual maturation. Finally, using a multi-omic data integration workflow supported by experimental validation, we show that the inhibitory nature of these cells is driven by specifically expressed secretory factors that cooperate with the retinoic acid signaling pathway to transform the adipogenic state of CD142– ASPCs into a non-adipogenic, Areg-like one.

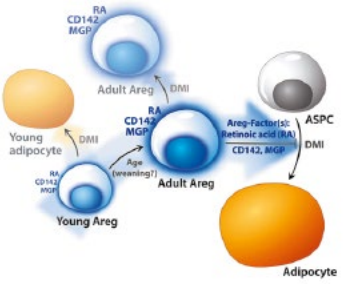

Fig. 10. *Graphical summary of the multi-omic and experimental characterization of mouse Areg-like cells (Zachara et al., 2021).*

Nevertheless, many questions remain. In the years to come, we therefore plan to invest substantial efforts to for example assess the physiological impact of Aregs on adipose biology, ideally by developing an Areg-specific ablation approach and monitoring the effect of eliminating these cells on adipose tissue growth and homeostasis as well as metabolic health in general under a control and high fat diet feeding regime. In addition, we would like to extend our studies to human adipose tissue, since the jury is still out with respect to whether Aregs are functionally conserved. Our initial findings pointed to the existence of a human CD142+ ASPC population with comparable molecular and cellular phenotypes to its mouse counterpart (Schwalie et al., 2018). Yet, while scRNA-seq-based analyses have revealed the existence of the “green” and “red” populations in human adipose stromal cells, these same analyses failed to resolve a clearly distinct transcriptomic signature that resembled that of mouse Aregs. Further analyses, involving more individuals / patients, also in different metabolic contexts, will however be required before firm conclusions can be drawn.

Engineering next-generation single cell phenomic technologies

DisCo: Deterministic scRNA-seq

Our scRNA-seq-based resolution of ASPC heterogeneity truly exposed us to the analytical power of single cell transcriptomics studies. However, it also revealed opportunities for technological improvements. For example, Aregs constitute a relatively minor proportion of all ASPCs (on average around 10%) and would as such be difficult to process as a standalone cell population with conventional scRNA-seq approaches, since these are all geared toward highest possible throughput. Indeed, little progress has so far been made to enable studies on small samples comprising <1,000 cells. To date, the Fluidigm C1 system, and other hydrodynamic trap-based methodologies, are the only mainstream systems able to process low input samples. Yet, cell size and shape capture biases have rendered these systems unpreferred for the processing of small heterogeneous samples, and thus of virtually any tissue. Hence, a wide array of precious biopsy samples and small cell populations or tissues are currently difficult to access by scRNA-seq. In response, such samples tend to be massively pooled (for example Zebrafish embryos, whole organism *C. elegans*, intestinal organoids etc.). This is rather counterproductive given that, despite having single cell-resolved transcriptomes, such sample pooling generally averages cell-type abundances and correlations, making the exploration of interindividual heterogeneity, particularly important in the context of developmental or clinical studies, impossible.

To address this technological gap, we developed a "no cell left behind" scRNA-seq platform, leveraging our expertise in multilayer microfluidics to engineer a DeterminIStic CO-encapsulation (DisCo) system for droplet-based scRNA-seq. Conventional Drop-seq allows for the straightforward processing of thousands of cells, which is why it is also so widely used. Its clear drawback is that the co-encapsulation process of one cell with one mRNA capturing bead in a droplet is uncontrolled, i.e. stochastic. This results in the generation of many droplets that either contain a cell or a bead but not both, making the process highly inefficient and incompatible with small samples since <20% of processed cells tend to be captured. In other words, >80% of the cells tend to be inevitably lost, which necessitates inputs of at least several thousands of cells. To address this im-

portant issue and thus to make Drop-seq compatible with small cell input samples, we supplemented the canonical Drop-seq chip with a control layer using multilayer microfluidics. This control layer is used to coordinate cells and beads at the co-encapsulation point (**Fig. 11**). The coordination is controlled by a machine-vision system, operating on bright-field (BF) images obtained from a microscope camera (**Fig. 11A**). Thus, by combining machine-vision and multilayer microfluidics and building on earlier advances by our lab to optimize mRNA capturing bead collection (Biočanin et al., 2019) we now achieved for the first time, to our knowledge, full control of a two-particle co-encapsulation process: precise placement of one cell and one bead, encapsulation by dropleting on demand, and on-chip sorting of dedicated droplets. Importantly, by performing precise cell-capture efficiency measures, we were able to demonstrate that our DisCo platform now routinely achieves >80% cell processing efficiency on 100 cells and less. This is a significant improvement over all currently available scRNA-seq technologies, and assured that the technology itself is no longer a limiting factor when processing low cell input samples. Rather, we believe that the most inefficient step in any experimental scRNA-seq setup is now the dissociation efficiency, which is unfortunately all too often ignored by the field as a whole. Much more efforts should therefore be devoted to improve the robust and comprehensive isolation of individual cells and assure a correct representation of all cell types in scRNA-seq samples.

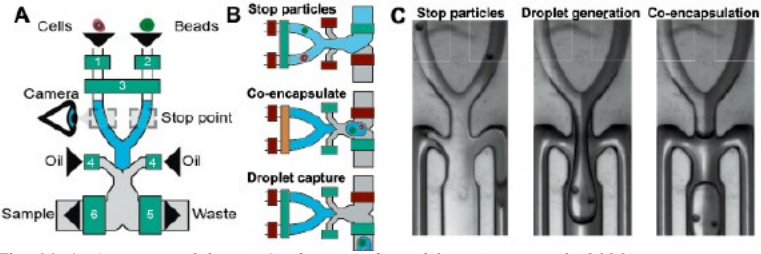

Fig. 11. A. *Overview of the DisCo device (adapted from (Bues et al., 2020)current microfluidics-based scRNA-seq technologies are limited to samples with large amounts of cells (> 1,000 cells), showing three inlet channels for cells, beads, and oil, and two outlets for waste and sample liquids. All inlets and outlets are augmented with "Quake valves" (green boxes): 1. cell, 2. bead, 3. dropleting, 4. oil, 5. waste, and 6. sample valves. The device is continuously monitored by a high-speed microscopy camera to detect and place particles at the Stop point. B. Illustration of the particle co-encapsulation process on the DisCo device (red: closed valve, green: open valve, orange: actuation for dropleting). C. The co-encapsulation process of two beads and droplet generation as observed on chip.*

Live-seq: transcriptomic recording of live cells

As exemplified by our studies on adipogenesis, most biological processes are inherently transient and dynamic, with cell states changing according to internal programs and/or external stimuli. It is thus critical to not only understand a cell's current state, but also how a cell arrived at that state, i.e. its molecular history. For example, ASPCs are able to differentiate in adipocytes, but as shown in **Fig. 9**, not all cells do so. Why this is the case remains a grand, outstanding question, but one of the hypotheses is that the pre-differentiation cell state may have been intrinsically refractory to, in this case, adipogenic stimuli. To test this hypothesis, one would ideally probe a cell's molecular state pre-differentiation and then track this same cell throughout adipogenesis with the aim of identifying factors that reflect either a cell's pro- or anti-adipogenic state. However, despite its obvious importance, revealing a cell's molecular history remains a great, outstanding technological challenge. Several approaches, from recombinase- and Crispr-based DNA editors, over live cell imaging, to single cell transcriptomics (scRNA-seq)-based trajectory inference have therefore been developed with the intent of exploring a cell's past. While constituting exciting advances, these methods have intrinsic limitations, including their ability to record only a couple of events per single cell rather than the whole transcriptome and thus by their reliance on prior knowledge of informative target genes or pathways. In addition, scRNA-seq-based trajectory models still need to be interpreted as statistical expectations rather than the real transition path of cells. This is because all current scRNA-seq assays depend on cell lysing to retrieve the respective transcriptome, which makes it impossible to link the individual cell to downstream molecular and phenotypic states.

To overcome the issue of lack of baseline data to predict cell responses or trajectories at the single cell level, in collaboration with Prof. Julia Vorholt's lab (ETHZ), we have recently established Live-seq, a single-cell transcriptome profiling approach that preserves cell viability during the extraction of RNA (Chen et al., 2021). Live-seq relies on fluidic force microscopy (FluidFM) to extract a cytoplasmic sample, coupled to a sensitive low-input RNA-seq strategy that our lab developed (**Fig. 12A-B**). Remarkably, benchmarking experiments evaluating the molecular impact of extracting up to 1 picoliter (pL) of cytoplasm from pre-adipocytes (IBA

cells) revealed few gene expression differences between unprobed cells (control IBA cells) and those 1h or 4h post Live-seq extraction (**Fig. 12C**). These observations suggest that Live-seq does not impose major perturbations on cellular function, suggesting, perhaps provocatively, that mammalian cells do not possess built-in defense mechanisms against such cytoplasmic sampling. This conclusion is further supported by the fact that probed cells generally show high post-extraction cell viability (>80%) and that they behave functionally similar to their untreated counterparts, as we, for example, have shown for LPS-treated RAW-G9-like macrophages (**Fig. 12D**). Thus, we believe that Live-seq opens a new avenue to link a cell's molecular state directly to its present and future phenotypic properties (**Fig. 12E**), providing an opportunity to acquire direct rather than inferred cell dynamics read-outs. To further explore Live-seq's capacity and provide a first proof-of-concept, we sampled 14 RAW-G9 cells a first time, then stimulated them with LPS, after which we sampled the same cells a second time (**Fig. 12E**), yielding two cells that passed the filtering criteria at both sampling points. While clearly limited in number, these two cells, each constituting two distinct points in the same t-SNE map (**Fig. 12F**), provide to our knowledge the first empirically determined, transcriptome-wide read-out of a cell's trajectory, transitioning from a pre-treatment to post-treatment (LPS) state (**Fig. 12B & F**). Moreover, we found that Live-seq and conventional scRNA-seq data of control and LPS-treated RAW-G9 cells could be properly integrated (**Fig. 12F**), which allowed us to unambiguously establish the correct trajectory of cells that were processed by conventional scRNA-seq.

In sum, we believe that Live-seq is orthogonal to any other scRNA-seq approach today in that it keeps cells alive while all other approaches do not. This, in turn, enables the transcriptome of the cell to be recorded prior to phenotyping and allows questions to be addressed that no other scRNA-seq method *directly* can. These include, as illustrated in (Chen et al., 2021), how molecular and cellular heterogeneity is established and what the actual (and not statistical) trajectory of cells is. We therefore anticipate that Live-seq has the potential to transform scRNA-seq from the current, end-point type assay into a real-time analysis workflow. For Live-seq to be widely relevant and transferable though, we will need to substantially increase its throughput and efficiency and reduce its overall experimental complexity. These constitute exciting technological challenges that we hope we will be able to address in my lab in the coming years.

Beyond cellular heterogeneity: understanding how regulatory variation induces phenotypic diversity

The genome is a remarkable molecular entity since it contains all the instructions to generate a multitude of different cell types that are derived from a single zygote. To understand development, it is therefore imperative to be able to decode these instructions, as I have explained above. Deciphering these same instructions is however also crucial to understand phenotypic variation, since it is now widely accepted that seemingly small divergences in this "regulatory grammar" are for a large part responsible for interindividual differences in complex traits such as height, but also in disease susceptibility. Indeed, well over 10 years of genome-wide studies have revealed that the majority of common trait or disease-associated genetic variants fall into non-coding, likely regulatory regions and affect transcriptional programs. Consequently, resolving how variation in regulatory sequences is translated into phenotypic variation at the molecular, cellular, or organismal level is of great biomedical importance. However, few studies have so far been able to mechanistically disentangle how regulatory variants contribute to human inter-individual variability. This is further illustrated by the newly launched International Common Disease Alliance (ICDA, http://icda.bio), aiming to address this "variant to function" challenge head-on. The aim here is to develop novel method-

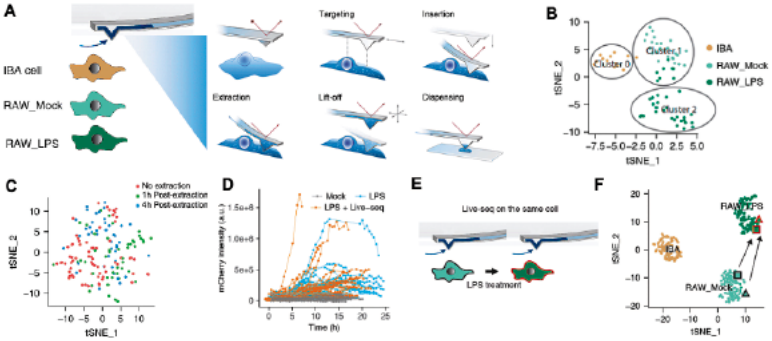

Fig. 12. A. *Live-seq workflow involving FluidFM and sensitive scRNA-seq to determine the molecular state of distinct live cell types and states (here, pre-adipocytes (IBA) versus mock- or LPS-treated macrophages* [*RAW-G9*]*).* ***B.*** *Unsupervised clustering of IBA and RAW-G9 Live-seq samples based on the top 500 variable genes as visualized in a tSNE plot.* ***C.*** *Unsupervised clustering of conventional scRNA-seq samples based on the top 500 variable genes as visualized in a tSNE plot. Shown in the plot are IBA control cells or IBA cells 1h, or 4h post-Live-seq extraction. The latter IBA cells do not show clearly distinct clustering, suggestive of high transcriptional similarity.* ***D.*** *RAW-G9 cells containing an mCherry reporter under the control of the* ***Tnf*** *promoter and GFP-tagged RELA (NFkB). mCherry intensity profiles of LPS-treated control RAW-G9 cells (blue) or those subjected to Live-seq sampling (orange). Mock-treated cells were used as negative control (grey). Such profiles can be derived since the FluidFM system is mounted on an optical microscope, which allows for monitoring cells in real time or in a time-lapse manner. Importantly, no striking behavioral differences were observed between the control and Live-seq-probed cell profiles.* ***E****. Sequential sampling procedure. The shape outline represents unstimulated (black) or LPS-stimulated cells (red).* ***F.*** *tSNE-based visualization of integrated scRNA-seq and Live-seq data, highlighting the transition of two sequentially sampled cells (triangle and square) from one state (ground) to another (LPS). The annotation of these cells is as described in (****B & E****). (Adapted from* [*Chen et al., 2021*]*).*

ologies and concepts that go well beyond *a priori* or even intuition-based knowledge of the gene(s) that likely mediate(s) the observed phenotypic variation, which, so far, is still the most common approach. While such gene-centred strategies greatly reduce the search space for causal variants, they also restrict the research scope to mostly gene-proximal regulatory elements (REs) such as promoters or introns.

To close this mechanistic gap, additional molecular traits, including TF binding, chromatin accessibility and state, are increasingly being assayed, again aiming to find causal variants by prioritizing variants in function of their location in functional genomic regions. While the integration of these additional molecular layers is challenging, it has nevertheless propelled the elucidation of rather convoluted molecular scenarios. A prime example involves obesity-associated variants that are located in the FTO gene. While initially hypothesized to affect the FTO gene itself, a whole palette of experimental and computational approaches has since demonstrated that the causal variant is indeed positioned in an FTO intron but affects the expression of two TF-coding genes (*IRX3* and *IRX5*). The latter genes are located more than 1 Mb downstream of the focal variant that disrupts the binding ability of yet another TF, ARID5B (Claussnitzer et al., 2014, 2015; Smemo et al., 2014). This appears to influence white fat cell function, providing a plausible link to excessive fat accumulation and obesity. Thus, this labour-intensive case study clearly emphasizes the need for novel approaches or concepts that could facilitate our ability to disentangle the role of regulatory variation in complex traits and disease susceptibility.

Another striking finding that emerged from recent, functional genomics studies is that the manner by which genetic variation impacts supposedly less convoluted phenotypes, such as gene regulatory processes, is also more complex than anticipated. We contributed to this important realization through a large-scale, integrative genomics study that my lab undertook in collaboration with Profs. Manolis Dermitzakis (University of Geneva), Alex Reymond and Nouria Hernandez (both at the University of Lausanne) (Kilpinen et al., 2013). The goal of our study was to quantify the allelic coordination among different molecular phenotypes to increase our knowledge of the chain of regulatory events leading to the transcriptional readout of a gene. We thereby aimed to also improve our under-

standing of to what extent genetic variation affects the chromatin landscape including TF binding profiles and chromatin mark enrichment. For this purpose, we performed ChIP of five chromatin marks (H3K4me1, H3K4me3, H3K27ac, H3K27me3, and H4K20me1), three transcription factors (TFs) (TFIIB, PU.1, and MYC), and the second largest RNA polymerase II subunit RPB2 in lymphoblastoid cell lines (LCLs) from two parent-offspring trios. A subset of the ChIP assays was additionally performed in eight additional unrelated individuals. In addition, all 14 individuals were profiled for gene expression.

Using these data, we then investigated the allele-specificity and inter-assay coordination among different molecular phenotypes. We observed abundant allele-specific effects across all probed molecular phenotypes, with the proportion of significantly biased sites ranging from 5% in mRNA to 11-12% in TF data and 6-30% in chromatin marks. We assessed the degree of parental transmission of the allelic effects, a proxy for genetic influence on the assays, and discovered that effects of DNA sequence variation are largely transmitted from parents to children from TF binding through chromatin marks to transcription. Transmission of allelic chromatin mark effects appeared much more sensitive to context-dependent effects compared to TFs though, with strongest transmission seen at promoters (for H3K4me3) and known chromatin accessibility-affecting variants (for H3K4me1 and H3K27ac). Interestingly, we also observed highly coordinated allelic (local) and haplotypic (short- and long-range) behavior among molecular phenotypes at different functional elements of the genome, suggesting that TF binding, presence of histone modifications, and the transcriptional readout at these regions all operate within the same allelic framework. Therefore, genetic effects on chromatin marks are probably closely tied to the sequence context of a given functional element and thus manifested indirectly through TF binding, which we observed to be causatively affected by DNA sequence variants within binding motifs of the same or other TFs.

The latter observation is nicely illustrated in **Fig. 13A**, which shows that the impact (in terms of DNA binding affinity) of SNPs on TF motifs (here, that of PU.1) scales with the likelihood to observe significant allele-specific effects, suggesting that the SNP-mediated disruption of the TF motif is directly causal to the observed allele-specific binding of the respective

TF. Nevertheless, a striking finding that emerged from these analyses is that despite PU.1's strong binding preference for its own motif (**Fig. 13A**), only a minority of variable PU.1 DNA binding events could be attributed to variants that disrupted the PU.1 motif (**Fig. 13B**). In other words, for the majority of PU.1 binding (70%) events that exhibit allelic DNA binding bias, the PU.1 motif is intact, suggesting that other factors / mechanisms are influencing PU.1 DNA binding. In our search to uncover these mechanisms, we found that for a small portion of these focal motif-independent variable PU.1 DNA binding events, disruptions in other TF motifs may also be responsible for changes in PU.1 DNA binding, implying cooperative effects between TFs. Specifically, a scan for allelic binding cooperativity within individuals identified four motifs that show a significant correlation between motif covariance and allele-specific PU.1 binding (**Fig. 13C**), collectively explaining another portion (7.5%) of all detected significant allele-specific PU.1 binding sites.

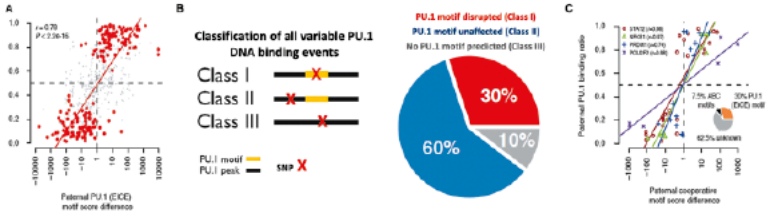

Fig. 13. *Genome-wide analyses of allele-specific PU.1 binding.* ***(A)*** *PU.1 motif score changes are predictive of allele-specific PU.1 binding. Ratio between paternal and maternal PU.1 PWM scores (x-axis) and fraction of reads mapping to the paternal allele (y-axis) (Red, significant sites; gray, non-significant).* ***(B)*** *Classification and proportion of all variable PU.1 DNA binding events according to three listed scenarios.* ***(C)*** *SNPs in co-associated TF binding sites are predictive of allele-specific PU.1 binding (5% FDR). (Adapted from [Kilpinen et al., 2013]).*

Together, our analyses support the notion of TFs being the primary mediators of sequence-specific regulation of gene expression programs, while chromatin marks are more prone to stochastic, possibly transient effects (e.g. due to environmental triggers) and likely reflect, rather than define, coordinated regulatory interactions. Our study also raised important, fundamental questions including how the uncovered, often long-range molecular coordination is genomically organized and specifically,

why and how only the minority of genetically variable TF binding events can be explained by sequence differences in the respective binding sites. To address these questions, we expanded our initial study by performing genome-wide profiling of PU.1 and three chromatin marks (H3K4me3, H3K4me1 and H3K27ac) as well as gene expression (RNA-seq) in lymphoblastoid cell lines of 47 individuals whose genomes were well characterized. Integrating these data, we observed strong quantitative co-variation between TF binding and chromatin mark levels at distinct regulatory regions, revealing a fine-grained molecular modularity of the genome that we newly defined as sub-megabase scale "variable chromatin modules (VCMs)" (**Fig. 14**) (Waszak et al., 2015). In addition, we found that the activity level of most VCMs can be captured by a single quantity value, which suggests that the molecular processes within each VCM (TF binding, chromatin mark enrichment, and gene expression) are all subject to one or few highly penetrant causal events. As such, VCMs may provide a conceptual framework as to why most regulatory variation is independent of local genetic variation. Indeed, we now hypothe-

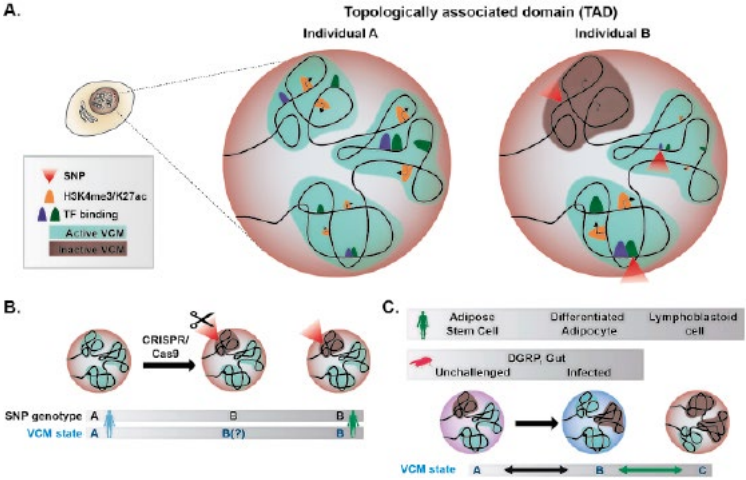

Fig. 14. *VCMs are genomic modules at sub-TAD, i.e. sub-Mb scale that capture coordinated variation in TF binding, chromatin state, and gene expression. We hypothesize that VCM activity is driven by TF-DNA interactions and that the perturbation of a single or few TF binding events may influence the molecular state of the entire VCM (taken from* [*Deplancke et al., 2016*]*).*

size that such regulatory variation is in large part driven by the activity state of the VCM in which the respective molecular phenotypes such as TF binding or chromatin mark enrichment are embedded.

This in turn shifts the question toward which genetic or molecular factors control the activity state of a VCM? This fundamental question is central to the studies that are currently ongoing in my lab and that leverage several important clues that have recently been reported. These include i) that genetic variation can influence chromatin accessibility independent of expression change and ii) that genetic variation in some, but not all regulatory elements within a certain locus influences the activity profile of all other molecular phenotypes in that locus. The latter findings indicate that regulatory elements not only undergo genotype-specific changes in accessibility, as presumably mediated by the gain or loss of cell type-specific TF binding, but that they may also be subjected to a certain regulatory hierarchy, involving both "lead" and "dependent" regulatory elements. It is tempting to speculate that this regulatory hierarchy is conceptually linked to the VCM principle. Deciphering which molecular features distinguish lead from dependent regulatory elements and thus how regulatory hierarchies are established across cell types and states are therefore among the main research themes in my lab. This is because resolving these hierarchies may provide us with key insights into how VCMs are established across systems, which in turn may prove instrumental to understand the contribution of genomic variation to molecular, cellular, and organismal diversity. The underlying rationale thereby is that we hypothesize that genetic variants that affect VCM activity or underlying regulatory hierarchies tend to have long-range impact on surrounding molecular phenotypes, increasing the likelihood that they induce downstream molecular and cellular effects, especially if they overlap with variants that impact gene expression. Thus, among the many variants that have so far been associated with specific traits or diseases, we would like to argue that those that affect VCM activity should be prioritized for further characterization given their anticipated, significant impact on local regulatory networks, which renders them attractive candidate drivers of phenotypic variation. We would thereby be able to truly exploit the coordinated molecular nature of a VCM of interest (e.g. linked to a variant with clear phenotypic / disease impact) for uncovering the flow of regu-

latory information, from causal nucleotides over gene(s) to phenotype, thus providing unprecedented insights into the molecular mechanisms driving phenotypic variation.

Conclusions and outlook

This year (2021) marked the 20th anniversary of the release of the human genome sequence. In these two decades and aided by immense technological and computational advances to which my lab also made a few contributions, tremendous progress has been made in our understanding of how our genome generates this stunning cellular and phenotypic diversity. The real challenge ahead will now be to move beyond the type of descriptive work that allowed us to catalogue the principal elements in our genome (genes, regulatory sequences etc.) to a mechanistic understanding of how all these elements work together to create a functional cell or organism. As such, I believe that the challenge is no longer in generating omic data, but in the way such data will be analysed and integrated with other datasets to decipher and even reverse engineer complex biological processes or systems.

For example, thanks to large-scale efforts such as ENCODE and contributions by numerous other groups, we now have great oversight of which TFs are encoded by a genome, which are the regulatory elements at which these TFs are active and in which chromatin context these elements are embedded in function of cell state or type. However, this does not suffice to understand the underlying regulatory code, as we are still unable today to robustly engineer from scratch a synthetic DNA sequence that allows us to induce a regulatory activity in a specified cell type and at a pre-defined time and level. Consequently, we also still have great trouble in predicting the effects of regulatory variants on gene expression, let alone at the cellular or organismal level. This is further exacerbated by the fact that our work on VCMs has taught us that regulatory elements operate in coordinated fashion and abide to certain hierarchies. This is why we need to not only better understand the impact of genetic perturbations locally, but also from a more integrative, network or regulatory module perspective. Only then do I believe will we be able to truly grasp how regulatory networks operate to define cell states and to enact cell fate transitions

such as from a mesenchymal stromal cell to a mature adipocyte, and what might be the molecular and phenotypic consequences when such networks are genetically perturbed. With the truly exciting advances in single cell omics, whose contributions, I anticipate, will reach well beyond the field of regulatory genomics into practically every aspect of biology and medicine, we now have an unprecedented opportunity to define these cell states / types with very high precision. This is an absolute pre-requisite in our quest to link regulatory networks to cell identities. In addition, novel deep learning approaches that now even consider genomic architecture are increasingly emerging and appear to rapidly improve our ability to infer gene expression from DNA sequence. These advances offer hope that we will be able to extract more general regulatory rules that then can be tested, for example using forward genetics approaches such as CRISPR-based sequence interrogations. The ultimate goal is then to apply these rules on each person's genome, catalysing the long-anticipated transition of clinical practice toward precision medicine.

Acknowledgements

I would like to sincerely thank the Professor Dr. Max Cloëtta Foundation for appreciating and honouring the type of interdisciplinary work that my lab is involved in by this very prominent award. My scientific journey could never have reached its current destination without the significant support and guidance of numerous people: my Master's thesis supervisor Prof. Willy Verstraete (Ghent University), who triggered my scientific curiosity, my PhD thesis advisor Prof. Rex Gaskins (University of Illinois), who taught me scientific rigor and persistence, and my postdoctoral advisors Profs. Marc Vidal and Marian Walhout (Harvard Medical School and University of Massachusetts Medical School) who introduced me to the exciting worlds of regulatory genomics and systems biology. I further would like to sincerely thank my initial Dean at the EPFL School of Life Sciences, Prof. Didier Trono, as well as Institute of Bio-engineering Director, Prof. Jeff Hubbell, who welcomed me with open arms and who were willing to invest in my lab's success, while establishing a fantastic scientific and administrative environment that continues to inspire and catalyse progress. A big thanks as well to all of my local and inter-

national collaborators whose vital input and support has broken research barriers and my deepest gratitude to all my former and current lab members, who are the real stars in this story. Indeed, I learned probably more from you than you did from me, so this prize really belongs to all of you as well. I would also like to acknowledge all my funding sources, including the EPFL, Swiss National Science Foundation, Chan Zuckerberg Foundation, EMBO, European Union Research Programs, HFSP, Inno-Suisse, Krebsliga, Precision Health and related Technologies Program, and SystemsX. Finally, I would like to thank my friends, family/parents and especially wife, Nele, and two sons, Jasper and Jonas, whose unconditional love and trust allowed me to remain level-headed and keep perspective. "If I know what love is, it is because of you." (Herman Hesse)

References

Alpern, D., Gardeux, V., Russeil, J., Mangeat, B., Meireles-Filho, A.C.A., Breysse, R., Hacker, D., and Deplancke, B. (2019). BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol. *20*, 71.

Bevers, R.P.J., Litovchenko, M., Kapopoulou, A., Braman, V.S., Robinson, M.R., Auwerx, J., Hollis, B., and Deplancke, B. (2019). Mitochondrial haplotypes affect metabolic phenotypes in the Drosophila Genetic Reference Panel. Nat Metab *1*, 1226–1242.

Bioč anin, M., Bues, J., Dainese, R., Amstad, E., and Deplancke, B. (2019). Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip. Lab Chip *19*, 1610–1620.

Bou Sleiman, M., Frochaux, M.V., Andreani, T., Osman, D., Guigo, R., and Deplancke, B. (2020). Enteric infection induces Lark-mediated intron retention at the 5' end of Drosophila genes. Genome Biology *21*, 4.

Bou Sleiman, M.S., Osman, D., Massouras, A., Hoffmann, A.A., Lemaitre, B., and Deplancke, B. (2015). Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. Nature Communications *6*, 7829.

Bues, J., Bioč anin, M., Pezoldt, J., Dainese, R., Chrisnandy, A., Rezakhani, S., Saelens, W., Gupta, R., Russeil, J., Saeys, Y., et al. (2020). Deterministic scRNA-seq of individual intestinal organoids reveals new subtypes and coexisting distinct stem cell pools. BioRxiv 2020.05.19.103812.

Chen, W., Schwalie, P.C., Pankevich, E.V., Gubelmann, C., Raghav, S.K., Dainese, R., Cassano, M., Imbeault, M., Jang, S.M., Russeil, J., et al. (2019). ZFP30 promotes adipogenesis through the KAP1-mediated activation of a retrotransposon-derived Pparg2 enhancer. Nat Commun *10*, 1–16.

Chen, W., Guillaume-Gentil, O., Dainese, R., Rainer, P.Y., Zachara, M., Gäbelein, C.G., Vorholt, J.A., and Deplancke, B. (2021). Genome-wide molecular recording using Liveseq. BioRxiv 2021.03.24.436752.

Claussnitzer, M., Dankel, S.N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., et al. (2014). Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell *156*, 343–358.

Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. New England Journal of Medicine *373*, 895–907.

Dainese, R., Gardeux, V., Llimos, G., Alpern, D., Jiang, J.Y., Meireles-Filho, A.C.A., and Deplancke, B. (2020). A parallelized, automated platform enabling individual or sequential ChIP of histone marks and transcription factors. PNAS *117*, 13828–13838.

Deplancke, B. (2009). Experimental advances in the characterization of metazoan gene regulatory networks. Briefings in Functional Genomics and Proteomics *8*, 12–27.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J. (2004). A gateway-compatible yeast one-hybrid system. Genome Res *14*, 2093–2101.

Deplancke, B., Alpern, D., and Gardeux, V. (2016). The Genetics of Transcription Factor DNA Binding Variation. Cell *166*, 538–554.

Ferrero, R., Rainer, P., and Deplancke, B. (2020). Toward a Consensus View of Mammalian Adipocyte Stem and Progenitor Cell Heterogeneity. Trends in Cell Biology *30*, 937–950.

Frochaux, M.V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., Braman, V.S., Andreani, T., Osman, D., and Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila genetic reference panel. Genome Biol *21*, 6.

Gubelmann, C., Waszak, S.M., Isakova, A., Holcombe, W., Hens, K., Iagovitina, A., Feuz, J.D., Raghav, S.K., Simicevic, J., and Deplancke, B. (2013). A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. Molecular Systems Biology *9*, 682.

Gubelmann, C., Schwalie, P.C., Raghav, S.K., Roder, E., Delessa, T., Kiehlmann, E.,

Waszak, S.M., Corsinotti, A., Udin, G., Holcombe, W., et al. (2014). Identification of the transcription factor ZEB1 as a central component of the adipogenic gene regulatory network. ELife *3*, e03346.

Hens, K., Feuz, J.-D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E., and Deplancke, B. (2011). Automated protein-DNA interaction screening of Drosophila regulatory elements. Nat Meth *8*, 1065–1070.

Isakova, A., Berset, Y., Hatzimanikatis, V., and Deplancke, B. (2016). Quantification of Cooperativity in Heterodimer-DNA Binding Improves the Accuracy of Binding Specificity Models. Journal of Biological Chemistry *291*, 10293–10306.

Isakova, A., Groux, R., Imbeault, M., Rainer, P., Alpern, D., Dainese, R., Ambrosini, G., Trono, D., Bucher, P., and Deplancke, B. (2017). SMiLE-seq identifies binding motifs of single and dimeric transcription factors. Nature Methods *14*, 316–322.

Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., et al. (2013). Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. Science *342*, 744–747.

King, M., and Wilson, A. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.

Litovchenko, M., Meireles-Filho, A.C.A., Frochaux, M.V., Bevers, R.P.J., Prunotto, A., Anduaga, A.M., Hollis, B., Gardeux, V., Braman, V.S., Russeil, J.M.C., et al. (2021). Extensive tissue-specific expression variation and novel regulators underlying circadian behavior. Sci Adv *7*, eabc3781.

Pradhan, R.N., Bues, J.J., Gardeux, V., Schwalie, P.C., Alpern, D., Chen, W., Russeil, J., Raghav, S.K., and Deplancke, B. (2017a). Dissecting the brown adipogenic regulatory network using integrative genomics. Scientific Reports *7*, 42130.

Pradhan, R.N., Zachara, M., and Deplancke, B. (2017b). A systems perspective on brown adipogenesis and metabolic activation: Systems perspective on brown adipogenesis. Obesity Reviews *18*, 65–81.

Raghav, S.K., Waszak, S.M., Krier, I., Gubelmann, C., Isakova, A., Mikkelsen, T.S., and Deplancke, B. (2012). Integrative Genomics Identifies the Corepressor SMRT as a Gatekeeper of Adipogenesis through the Transcription Factors C/EBPβ and KAISO. Molecular Cell *46*, 335–350.

Schwalie, P.C., Dong, H., Zachara, M., Russeil, J., Alpern, D., Akchiche, N., Caprara, C., Sun, W., Schlaudraff, K.-U., Soldati, G., et al. (2018). A stromal cell population that inhibits adipogenesis in mammalian fat depots. Nature *559*, 103–108.

Simicevic, J., Schmid, A.W., Gilardoni, P.A., Zoller, B., Raghav, S.K., Krier, I., Gubelmann, C., Lisacek, F., Naef, F., Moniatte, M., et al. (2013). Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. Nat Meth *10*, 570–576.

Simicevic, J., Moniatte, M., Hamelin, R., Ahrné, E., and Deplancke, B. (2015). A mammalian transcription factor-specific peptide repository for targeted proteomics. PROTEOMICS *15*, 752–756.

Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.

Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A., et al. (2015). Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. Cell *162*, 1039–1050.

Zachara, M., Rainer, P.Y., Russeil, J.M., Hashimi, H., Alpern, D., Ferrero, R., Litovchenko, M., and Deplancke, B. (2021). Mammalian adipogenesis regulators (Aregs) exhibit robust non- and anti-adipogenic properties that arise with age and involve retinoic acid signalling. BioRxiv 2021.02.24.432431.

Zwick, R.K., Guerrero-Juarez, C.F., Horsley, V., and Plikus, M.V. (2018). Anatomical, Physiological, and Functional Diversity of Adipose Tissue. Cell Metabolism *27*, 68–83.