The elucidation of the gene regulatory networks underlying terminal adipogenesis

A summary description of the project objectives:

This project aims to improve our understanding regarding the structural and dynamic properties of gene regulatory networks (GRNs) underlying distinct biological processes. As a model system, we originally chose microglia because they constitute a readily accessible and relatively homogeneous cellular system in which the transcriptional mechanisms underlying its biological function are currently poorly understood. However and as clearly stipulated in the first period report, we switched our focus to characterizing the GRNs underlying fat cell differentiation because of practical and ethical issues (i.e. many mice would have to be sacrificed to generate a primary microglial cell culture). We reasoned that this is a valid alternative because studying adipogenesis also has clear, medical relevance, because well-established *in vitro* adipogenic models are available, and finally because adipogenesis occurs through a cascade of gene expression events, which are mediated by GRNs. Elucidating their content and dynamic properties is therefore essential. Thus, although the specific biological system was altered (and approved by the Commission), the overall scientific philosophy and structure of the grant remained intact, i.e. to determine the temporal gene expression profiles of transcription factors (TFs) during a biological process (now fat cell differentiation), to locate regulatory elements involved in this process, to map interactions between TFs and uncovered regulatory elements, and finally to identify and characterize critical adipogenic TFs. To perform these experiments, we proposed to develop a mouse-specific, gene-centered protein-DNA interaction screening method, and to implement genome-wide techniques to validate some of the detected interactions. Together, we reasoned that the obtained knowledge may provide an experimental and data framework for future modeling efforts to make predictions on how adipogenic GRNs behave under different physiological or pathological conditions, or on how to manipulate these networks such that excessive adipocyte accumulation and its deleterious consequences can be prevented or suppressed. Thus, given the well-established relationship between excess fat mass and the metabolic syndrome as well as cancer, this work may not only support important technical advances and generate resources and data of interest to the scientific community at large, it may also significantly increase our understanding of the molecular mechanisms underlying these pathologies.

A description of the work performed since the beginning of the project:

1. Establishment of a high-throughput screening method to detect mouse TF-DNA interactions
2. Mapping of TF-coding gene transcription profiles during adipogenesis
3. Generation of a gene- or transcript-specific qPCR primer database
4. Large-scale overexpression screen to identify adipogenic TFs
5. Development of a microfluidics-based method to map TF binding sites at high resolution
6. Mapping of repressor-associated regulatory elements
7. Characterization of a select number of TFs in terms of their gene regulatory function during adipogenesis

A description of the main results achieved so far:

1) High-throughput, mouse protein-DNA interaction detection method. To enable protein-DNA interaction mapping in the mouse and thus to allow the generation of a *core* GRN encompassing all possible regulatory interactions between adipogenic TFs and their target genes, we generated over the course of this project a comprehensive TF open reading frame (ORF) clone library. Out of 1576 possible TFs, we were so far able to clone 1050 and fully sequence-verify ~800. The latter was performed using a novel, automated decision full-length clone sequence verification approach using high-throughput sequencing data, which incorporates a novel sequence assembly algorithm that was developed by our group and published in *Nature Methods* [1]. A paper describing a user-friendly web interface to apply this approach was published in *Nucleic Acids Research* [2]. To allow the manual screening of over 1000 individual TFs versus each regulatory element, we established a fully automated yeast transformation robotic platform as published in *Nature Methods* as well [3].

2-4) Identification of adipogenic TFs. To identify relevant adipogenic TFs and monitor their transcription and expression profiles, we first analyzed genome-wide DNA-binding data of RNA polymerase II (RNAPII) at TF-coding genes, allowing us to cluster TF genes according to their transcriptional activity. We detected seven distinct transcriptional clusters used fuzzy c-means clustering, containing ~900 distinct TFs and thus revealing for the first

time the identity of and possible hierarchy among adipogenic TFs. To examine in how much these transcriptional profiles correlate with gene expression profiles, we performed quantitative RT-PCR (qPCR) assays on the 20 most representative members of each transcription cluster using primers retrieved from our gene- or transcript-specific qPCR primer database GETPrime as recently published in *Database* [4]. We found an overall concordance between mRNA and RNA pol II occupancy levels and we are currently examining the biological significance of each cluster by large-scale phenotypic screens as will be discussed in more detail below.

5) Development of a microfluidics-based method, MARE, to map TF binding sites at high resolution. MARE (i.e. **M**itomi-based **A**nalysis of **R**egulatory **E**lements) is best compared to 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. However, MARE allows conducting more than 700 of these EMSA-like assays at once on one microfluidic chip in relatively straight-forward and cost-effective fashion, enabling the fragmentation of long regulatory DNA sequences in small DNA elements, each of which can then be tested for binding to a specific set of TFs. As such, a relative DNA occupancy landscape can be generated for each TF over the length of the respective regulatory element and regions with highest occupancy then correspond to likely binding sites. Thus, MARE allows the simultaneous validation and localization of protein-DNA interactions within regulatory elements, and does not require *a priori* binding site information as it in principle can also be used to derive DNA binding specificity data. The method was included in our recent *Nature Methods* publication [3].

6) Mapping of repressor-associated regulatory elements. Consistent with one of the main goals of the grant to explore regulatory elements within the mammalian genome, we performed an integrative genomics study of a co-repressor, SMRT, which has a critical role in adipogenesis. The results of this study have recently been published in *Molecular Cell* [5], and allowed us to uncover novel gene regulatory mechanisms underlying adipogenesis.

7) **Characterization of a select number of TFs in terms of their gene regulatory function during adipogenesis.** We are currently generating stable cell lines for 20 TFs that were identified in our large-scale phenotypic screen as having either an activating or a repressing, regulatory function in adipogenesis. The aim is to perform genome-wide DNA binding studies on these TFs in fat cells, and as such to elucidate their role within the adipogenic GRN.

The expected final results and their potential impact and use

Our efforts are significantly expanding our knowledge of the GRNs underlying adipogenesis, which in turn will contribute to our overall understanding of the molecular mechanisms mediating this key differentiation process. As such, we hope to shed new light on possible therapeutic targets or strategies which may be useful to combat the ever expanding obesity epidemic, which is linked to several metabolic syndrome symptoms such as type 2 diabetes, hypertension, dyslipidemia, and even cancer. For example, pending the successful *in vivo* validation of key factors or interactions in future projects, we could envision that our adipogenic GRN models may allow us to make predictions on how these GRNs will behave under distinct physiological or pathological conditions, or on how to manipulate these networks such that excessive adipocyte accumulation and its deleterious consequences can be prevented or suppressed.

Next to its biological and medical relevance, this project also generated resources such as the mouse TF ORF clone library as well as novel experimental frameworks such as the high-throughput yeast one-hybrid and MARE assays. These will undoubtedly be useful for other researchers aiming to dissect regulatory networks underlying biological processes of their interest, as is reflected by the fact that they have been published in high profile journals such as *Nature Methods*.

References

1 Massouras, A. *et al.* Primer-initiated sequence synthesis to detect and assemble structural variants. *Nat Meth* **7**, 485-486, (2010).

2 Massouras, A., Decouttere, F., Hens, K. & Deplancke, B. WebPrInSeS: automated full-length clone sequence identification and verification using high-throughput sequencing data. *Nucl. Acids Res.*, gkq431, (2010).

3 Hens, K. *et al.* Automated protein-DNA interaction screening of Drosophila regulatory elements. *Nat Meth* **8**, 1065-1070, (2011).

4 Gubelmann, C. *et al.* GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. *Database* **2011**, (2011).

5 Raghav, Sunil K. *et al.* Integrative Genomics Identifies the Corepressor SMRT as a Gatekeeper of Adipogenesis through the Transcription Factors C/EBPβ and KAISO. *Molecular Cell* **46**, 335-350, (2012).