universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Testing combinatorial transcription factor activities using multidimensional recruitment assays“

verfasst von / submitted by

Franziska Reiter, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Dr. Alexander Stark

Acknowledgements

I would like to thank Alex for taking me into the Lab and giving me the opportunity to do cool science. For always being a great mentor, supporting me and giving me more opportunities than I probably deserve.

Thanks to the whole Stark lab. You guys make coming to work fun.
In particular, I want to thank Gerald, who set up the luciferase protocol in a way that made it almost no work, who helped me with everything when I first joined the lab and never said no to discussing a problem even if he had an important doctor exam the next day or was at the other end of the world.
Thanks to Ashley, Vanja and Mamduh for giving the promptest answers to a computer noob like me.
Thanks to Michi and Martina for making working in the lab so incredibly easy.
Thanks to Lukasz and all the people working on Human STARR-seq for showing me that science can also be hard ;).
Thanks to Christoph, Felix, Cosmas and Lukas for cool discussions, delicious dinners and not taking me too seriously.
Thanks to Rui for having a joke every time I need to laugh.

Table of contents

Zusammenfassung

Höhere eukaryotische Organismen bestehen aus einer Vielzahl von verschiedenen Zelltypen. All diese Zelltypen besitzen die gleiche Erbinformation und entstehen während der Embryonalentwicklung aus einer einzigen Zelle. Eine der interessantesten Fragen der Biologie zur Zeit ist, wie es möglich ist, dass eine so große Komplexität zustande kommen kann, obwohl jede Zelle das gleiche Genom innewohnt. Dies wird zu einem großen Teil durch die transkriptionelle Regulation von Genen erreicht. Enhancer sind cis-regulatorische Sequenzen die die korrekte zeitliche und räumliche Expression von Genen sicherstellen. Transkriptionsfaktoren binden an kurze Sequenzmotive im Enhancer und lesen somit die regulatorische Information die dort kodiert ist.

Transkriptionsfaktoren sind essenziell für Enhancer Funktion. Mutiert man die Bindestellen für einen Transkriptionsfaktor in einem Enhancer so ist dieser nicht mehr in der Lage seine regulatorischen Funktionen korrekt auszuführen. Umgekehrt verliert der Enhancer auch seine Aktivität wenn der Transkriptionsfaktor nicht vorhanden ist. Dies deutet darauf hin, dass einzelne Transkriptionsfaktoren nicht ausreichend sind um einen Enhancer zu aktivieren, sondern dass ein Kollektiv von Transkritionsfaktoren zusammenkommen muss um die korrekte Aktivität sicherzustellen.

In dieser Arbeit beschreiben wir einen Assay der es uns erlaubt mehrere Transkriptionsfaktoren an einen transkriptionellen Reporter zu rekrutieren, indem diese an verschiedene DNA Bindedomänen fusioniert warden die an distinkte DNA-Sequenzen binden. Damit ist es uns möglich deren cooperative Aktivität zu untersuchen. Wir verwendeten kontext-spezifische Transkriptionsfaktoren um gezielt nach deren Partnerfaktoren zu suchen. Wir testeten 476 Drosophila melanogaster Transkriptionsfaktoren und fanden 42 cooperative Paare. Diese Paare bestätigten sich in zwei Kontrollexperimenten. Keines dieser Paare ist jedoch ausreichend für transkriptionelle Aktivierung wenn man sie aus dem Enhancer Kontext heraußnimmt, und in einem synthetischen Kontext rekrutieret der nur die Erkennungssequenzen der DNA Bindedomänen enthält. Aus diesem Grund testeten wir Tripletts von Transkriptionsfaktoren, sowohl in einem Enhnacer Kontext als auch im synthetischen Kontext, und fanden cooperative Transkriptionsfaktoren in beiden Experimenten. Die

Kooperativität der Transkriptionsfaktoren wurde verstärkt wenn wir die natürliche Anordnung der Motive beibehielten.

Abstract

The temporal and spatial expression of genes is regulated by transcription factors (TFs) that bind to enhancer regions in a combinatorial fashion. Even though we know the identity of many TFs and the genes they regulate, it is unclear how exactly TFs control enhancer activity and gene transcription.

Here we probe the functional interdependencies of TFs and determine combinations of TFs that show synergistic activation. We co-recruit defined sets of TFs via different DNA-binding-domains (DBDs) to different positions within enhancer contexts. This multi-dimensional *enhancer complementation assay* revealed obligate combinatorial TFs and enabled the definition of pairs of TFs that strongly activate transcription when co-bound, even though each TF alone is inactive. Furthermore, we demonstrate that, even though both partner TFs are necessary for transcriptional activation, these cooperative TF pairs are not sufficient to reconstitute enhancer activity when co-recruited outside enhancer contexts. In contrast, enhancer function and reporter transcription can be achieved by recruiting three TFs simultaneously and is enhanced when they are recruited in an arrangement that reflects the binding site arrangement of an endogenous enhancer. The demonstration that TFs control transcription via combinations of (biochemically) distinct regulatory functions has important implications for our understanding of combinatorial enhancer control and gene expression (Reiter et al., 2016).

Introduction

Transcription in general

One of the most fascinating questions in biology is how the information in a single genome is differentially interpreted during development to give rise to the thousands of different cell types of a complex organism. Differential gene expression is controlled by genomic sequences called cis regulatory elements or enhancers. (Banerji et al., 1981; Levine, 2010; Yáñez-Cuna et al., 2013). Transcription factors (TFs) can bind to short sequence motifs within the enhancer and thereby read the information from the enhancer. They are regarded as the proteins responsible for determining the transcriptional output by recruiting transcriptional cofactors (Spitz and Furlong, 2012). The motifs they bind to are essential for enhancer function as mutating even one of them – or depleting the corresponding TF – can lead to a reduction or even a total loss of enhancer activity (Shlyueva et al., 2014; Yáñez-Cuna et al., 2014).

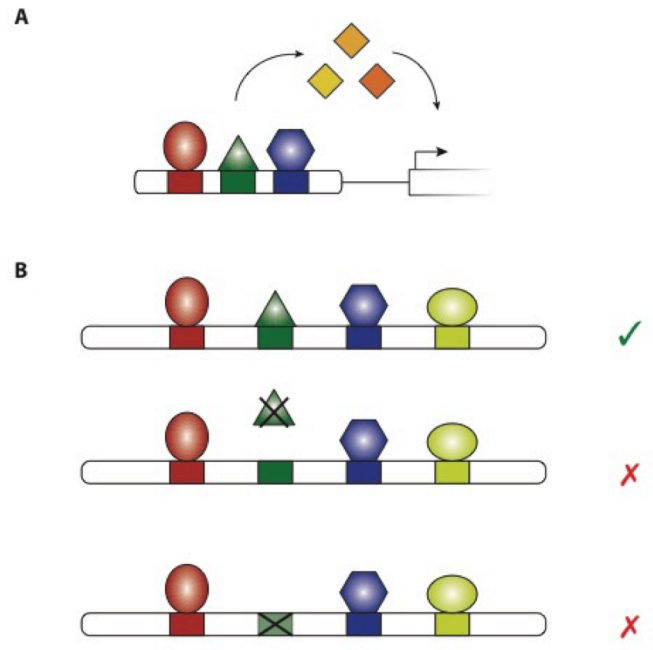

Fig 1: Transcription factors work together to activate gene transcription (a) Transcription factors bind to their sequence motifs in the enhancer and recruit cofactors, that activate transcription at the target gene. (b) Enhancers rely on TFs for their function. Depleting the TF or mutating their binding motif both result in loss of activity.

Transcription factors

The fact that the mutation of individual TF binding motifs and or the depletion of the corresponding TF disrupts enhancer function, suggests that enhancer activity relies on a defined set of TFs that function cooperatively in a strictly synergistic or combinatorial manner. This cooperative behavior of TFs has previously been observed during the dedifferentiation of fibroblasts into iPS cells, where a fixed set of TFs (Oct4, Sox2, cMyc and Klf4, together also called 'Yamanaka factors') (Takahashi and Yamanaka, 2006) is necessary and sufficient for this process. Even though some individual TFs are sufficient to induce a transcriptional program (MyoD overexpression is sufficient for induction of a skeletal muscle transcriptional program) (Weintraub et al., 1989), the complexity of a eukaryotic organism cannot be explained by a model where the development of every cell type is controlled by a single dedicated TF. It is therefore reasonable to assume that combinatorial regulation of TFs has evolved as a way to increase the possible number of different transcriptional programs that can be specified by a limited set of TFs. Cell signaling cascades for example utilize inducible TFs that cooperate with cell type specific TFs to activate their target genes in a controlled fashion (Heinz et al., 2010; Reiter et al., 2017). For example, the Drosophila *even-skipped* muscle and heart enhancer responds to Ras signaling in a cell type specific manner by integrating this signal with the mesoderm-restricted TFs Twist and Tinman (Halfon et al., 2000).

Transcription factor cooperativity

Mechanistically, cooperativity between TFs can follow one of two models. The first model regards a TF as an entity with a certain activation capacity. By recruiting TFs to an enhancer, the activation potential rises additively with the number of TFs (Crocker et al., 2016). On the other hand, the second model considers TFs to have distinct molecular and biochemical functions. In this model, a specific combination of TFs has to come together to activate transcription (Reiter et al., 2017; Stampfel et al., 2015; Yáñez-Cuna et al., 2014). The two models are not mutually exclusive as some TFs and/or enhancers might work through one, while others utilize the other model. Hints for the existence of the second model of TF cooperativity comes from studying the TF binding motifs at enhancers. Enhancers often show enrichment of different types of motifs. Experiments with synthetic arrays of TF motifs have found

that heterotypic clusters of motifs confer stronger activation than homotypic clusters (Fiore and Cohen, 2016; Gertz et al., 2008; Smith et al., 2013). In addition the disruption of a single motif can lead to a loss of enhancer activity that is greater than an additive model would predict, suggesting that the motif is required not only for the activating strength but for the overall activity of the enhancer.

This is not only true for enhancers that are constitutively active in a certain cell type (Fig. 2a; (Yáñez-Cuna et al., 2014)), but also for enhancers induced by signaling cues. In the fruit fly Drosophila melanogaster for example, the steroid hormone ecdysone activates a set of enhancers by via the ecdysone receptor (EcR) TF. Disruption of the EcR binding site completely abolishes the enhancer's responsiveness to ecdysone. Interestingly, a second motif, the GATA motif, is also essential for the hormone-induced activity (Fig. 2b; (Shlyueva et al., 2014)). Since both motifs are needed for enhancer activity it suggests that the TFs bound to these motifs cooperate either directly or indirectly.

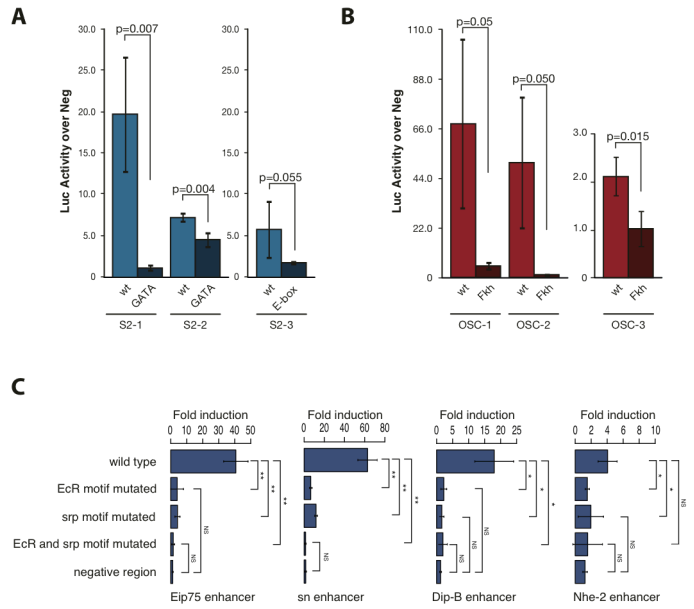

Fig 2: Enhancers contain several different motifs that can all be essential

Enhancers specific to different cell types rely on different sets of motifs. (a) S2 cell - (cells of embryonic tissue origin) specific enhancers rely on the GATA motif (b) OSC- (ovarian stem cells) specific enhancers rely on the Fkh (forkhead) motif. Mutating each one of several different motifs results in loss of enhancer activity (from Yañes-Cuna et al, 2014). (c) Fold induction after treatment with ecdysone. The enhancers' activities are lost after disruption of either EcR or GATA motifs, showing that both motifs are essential for enhancer activity (from Shlyueva et al, 2014).

These observations raise interesting questions about combinatorial enhancer control: do all TFs need partners, and if not, which ones do? There are well described cases of TFs that are sufficient to activate transcription on their own, however, most TFs do not. These TFs need additional

TFs at the enhancer to elicit a transcriptional response, but we know very little about how this cooperativity is achieved and which TFs cooperate with each other. For example, it is unclear if only very specific combinations of TFs are possible, or if classes of TFs exist that all work in a similar way and can substitute for each other. Furthermore, if such classes exist, do TFs only cooperate with TFs from the same class or do they require TFs from a different class and with potentially different functions, or are even all activating TFs interchangeable.

TFs have distinct regulatory activities post-binding

In a recent publication, we found groups of functionally distinct TFs using an ectopic reporter assay (Stampfel et al., 2015). We assayed the regulatory activities of different TFs by tethering them to different enhancer contexts using the GAL4-UAS system. This decouples the DNA binding activity of the TFs from their post-binding regulatory activity by fusing them to the DNA-binding domain (DBD) of the yeast transcription factor GAL4. The GAL4-DBD specifically and strongly binds to its recognition motif called upstream activating sequence (UAS). We assayed the activities of 474 Drosophila TFs by recruiting each of them to 24 different enhancer contexts, ranging from a synthetic 4xUAS binding site array, to fully functional enhancer sequences where a single TF motif in an enhancer was mutated to a UAS site. The synthetic context assays the TFs of interest in the absence of any Drosophila regulatory sequence, measuring the individual activating potential of the TF. Functional enhancer contexts on the other hand provide additionally the TFs that bind to this enhancer, making it possible to study cooperative functions of the TF.

Consistent with the additive model of enhancer function, we found global activators capable of activating the reporter independently of the context they were recruited to. Complementary to the global activators we found global repressors capable of repressing any given context. However, there were also TFs that only very specifically activated certain enhancer contexts. These context-dependent TFs seem to rely on the other TFs present in the respective enhancer context. This is illustrated by their inability to activate a synthetic 4xUAS reporter, as opposed to global activators which can activate a synthetic context to approximately the same level as the full enhancer contexts. One group of context-dependent TFs specifically

activated ecdysone-responsive enhancers when recruited to the EcR motif position. These TFs form a distinct cluster based on their activation patterns over the 24 different contexts (Fig. 3a b). One of these TFs is helix-loop-helix protein 4C (HLH4C), which in this assay could replace the EcR and activate the enhancer in an ecdysone independent manner.

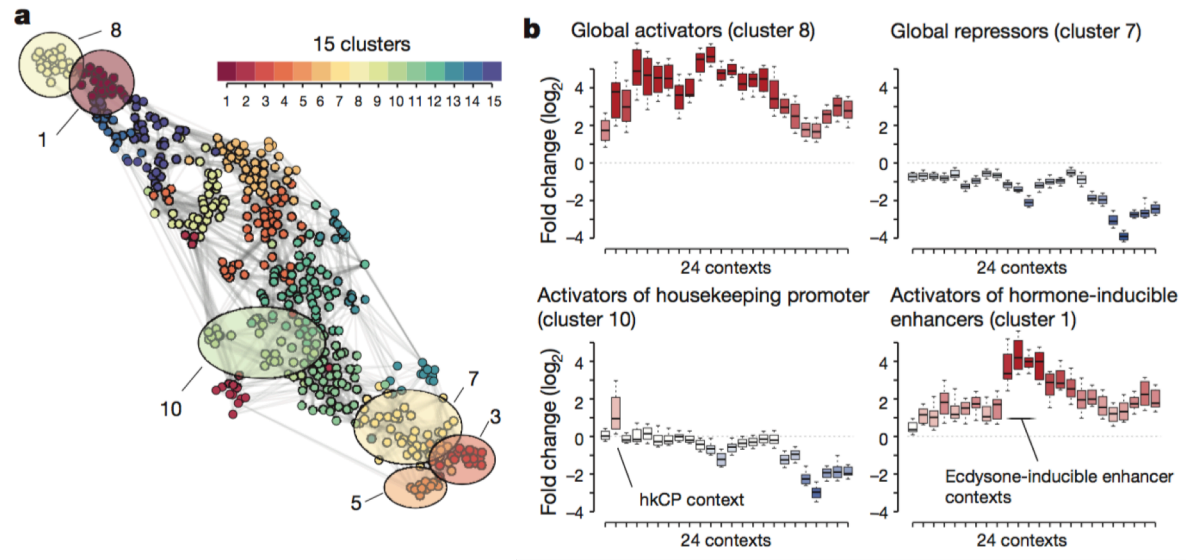

Fig 3: TFs form functionally distinct groups

(a) TFs fall into 15 clusters based on their activity across 24 different enhancer contexts. (b) normalized luciferase values of TFs from indicated clusters across the 24 contexts. (from Stampfel et al, 2015)

These context dependent TFs can also be described as obligately combinatorial, since they depend on the additional partner TFs present in the full enhancer context. Since these partner TFs are specific to the different enhancers, it could explain why some of the recruited TFs are only able to activate one type of enhancer context. However, even though we found TFs exhibiting this combinatorial behavior, this study did not reveal which additional TFs in the enhancer context they relied on i.e. what the required partner TFs were. To answer this question and find specific combinations of TFs where neither TF is able to activate transcription on its own, but that work together as a pair, we expanded the original assay to be able to recruit multiple TFs at the same time.

The observation of context dependent TFs together with the knowledge about essential motifs in the respective enhancer contexts, gives us a powerful tool to study cooperativity between pairs of TFs. TFs can be recruited to the position of either one of the essential motifs, either alone or together with a second TF. This lets us find obligatory TF pairs, that can only activate the reporter when recruited together. Since we have previous knowledge about the TFs that are able to substitute the EcR motif, we can specifically look for partners for these transcription factors, instead of testing all possible combinations of all TFs.

Aims of the thesis

Transcription factors have long been studied because of their essential role in transcriptional control. However, even though we know a lot about single TFs, less is known about how TFs work together to activate gene expression. We utilize two key properties of enhancers and TFs to study the cooperativity between TFs: 1) the fact that enhancers contain essential motifs, that can be replaced with the motifs of sequence specific DNA-binding domains which can be used to recruit TFs and 2) the existence of context dependent TFs. With this thesis we are trying to address the cooperativity between different TFs, which TFs can work together in a specific enhancer context and whether these combinations of TFs are sufficient for activation. The aims of this thesis are:

- Establish a multi-recruitment assay to measure the activating potential of different combinations of TFs
- Use this assay to screen for specific cooperative combinations of TFs
- Test the sufficiency of such combinations in a synthetic context

Results

DBDs with orthologous binding functionality for multi-dimensional recruitment

To be able to simultaneously recruit multiple TF to distinct sites in an enhancer, we cloned 11 additional DBDs from 4 different organisms. We indirectly tested each DBD's ability to selectively bind to its own recognition sequence by fusing it to the general activator p65 and measuring the activation of a luciferase reporter, consisting of the DBD's recognition sequence in front of a minimal core-promoter and a luciferase gene (Fig. 4a). Ideally a DBD would only bind to its own recognition sequence and therefore be able to activate only reporters containing this specific sequence. We recruited each DBD-p65 fusion protein to each of the 12 different

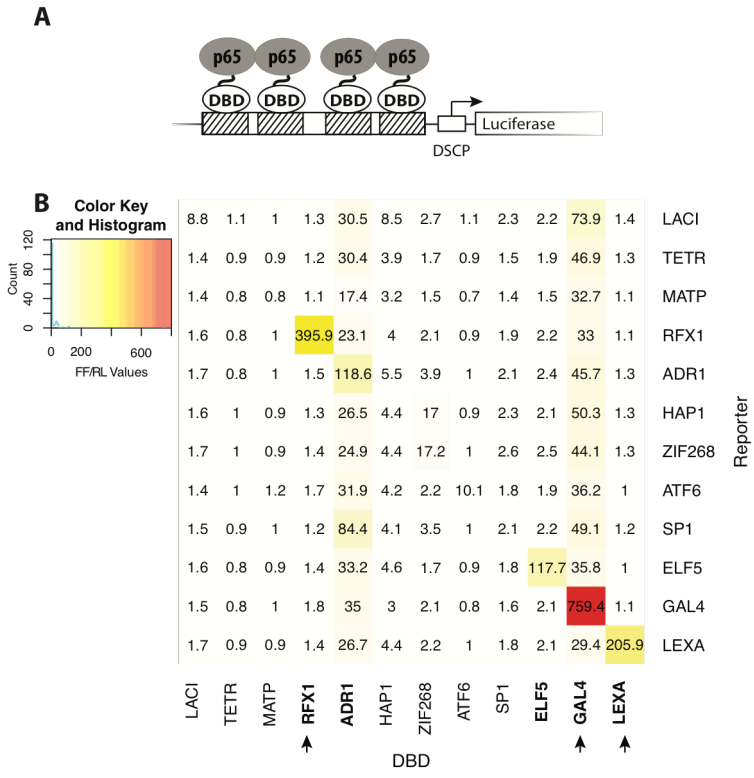

Fig 4: Different DBD-p65 fusion constructs activate the reporter with different strengths and specificities. Shown are 12 values (Firefly over Renilla, FF/RL) when recruiting the respective DBD-p65 fusion proteins to the indicated luciferase reporter constructs by co-transfection (see color legend). The DBDs of RFX1, ADR1, ELF5, GAL4 and LexA strongly and specifically activate reporters containing their recognition sequences. Indicated DBDs (arrows) were chosen for subsequent experiments.

reporters (one for each of the 11 candidate DBDs and one for the Gal4-DBD/UAS) to measure the activation of the correct reporter (i.e. the reporter containing the recognition sequence for the tested DBD) versus all other reporters (Fig. 4b). We did this by co-transfecting a plasmid containing the DBD-p65 fusion together with a plasmid containing the binding site and a reporter, and then assessing the transcriptional activity in dual luciferase assays across 4 biological replicates with independent transfections. This identified 4 additional DBDs that strongly and selectively activated their respective reporters (RFX1, ADR1, ELF5, and LEXA; GAL4 and ADR1 seem to have a more general DNA binding potential, but still activated their reporters more strongly than the others). Together with GAL4, this equipped us with 5 different DBDs of non-overlapping binding specificities that could be used in a multidimensional recruitment assay. We chose RFX1 and LexA as the additional DBDs next to GAL4 due to their ability to strongly activate the reporter when recruiting p65.

Reducing combinatorial complexity to design a screening set-up

To investigate cooperativity between TFs while avoiding the immense scale of performing 474x474 pairwise tests, we initially chose a single context dependent TF as a bait TF HLH4C, because of its strong context dependent behavior in the single-recruitment screen (Stampfel et al., 2015). Since we know that HLH4C only activates reporter transcription when recruited to certain contexts, we are able to specifically ask which other TFs in this context it might rely on by replacing an additional motif with a DBD binding site and testing which TFs can rescue activity. We kept HLH4C constant at the EcR motif position of the ecdysone-responsive enhancer Nhe2, and tested it against the 474 TFs in our library.

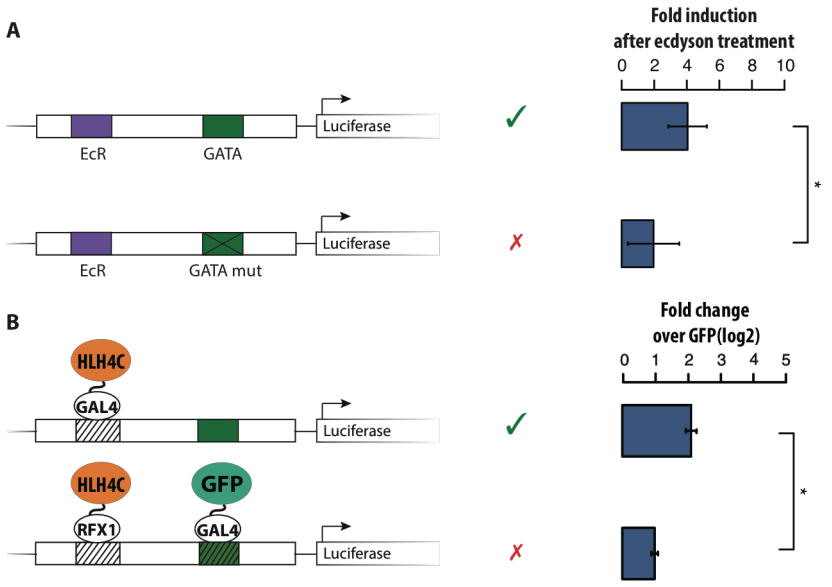

Fig 5: Recruitment of HLH4C at the EcR motif depends on the GATA motif for activation of the reporter. The induction of the wild type Nhe2 enhancer after ecdysone treatment is dependent on the GATA motif (top, from Shlyueva et al 2014). Similarly, when replacing the EcR motif with a UAS site and recruiting HLH4C, we only observe activation with an intact GATA motif (bottom).

For the double recruitment experiments we chose the DNA-binding domains RFX1 to recruit HLH4C and GAL4 to recruit each of 474 TFs (GAL4 was chosen because it had been used previously and therefore fusion constructs for all 474 TFs were already available). We changed the EcR motifs to a RFX1 binding site and replaced the GATA motifs with a UAS binding site. To check if in this scenario the same dependence on the GATA motifs holds true, we recruited HLH4C to the positions of the EcR sites and GFP to the positions of the GATA sites, mimicking the mutated GATA motifs (Fig. 5b). This produced results equivalent to mutating the GATA motif in the otherwise wild type enhancer after induction with ecdysone (Fig. 5a). As for the previous experiment (Shlyueva et al., 2014) activation of the reporter dropped by about half, suggesting that HLH4C, like the hormone-bound EcR depends on the TF bound to the GATA motif (Fig. 5a,b). This allowed us to recruit a second TF at the GATA motif and assay its ability to compensate for the mutated motif.

Assessing TF-TF cooperativity

Having established that the activating function of HLH4C recruited to the Nhe2-EcR-GATA context relies on the GATA motif and therefore presumably the TF bound to it, we wanted to test if we can also observe cooperative activation when recruiting a second TF there. Cooperative activation is given if neither TF can activate the reporter when recruited together with GFP, but only when co-recruited with another

cooperative TF or if the combined activities are more than additive. For the two TFs HLH4C and Ets96B this is the case (Fig 6a). Only the co-recruitment of the two TFs is able to activate the reporter, while their recruitment together with GFP does not lead to a strong activity.

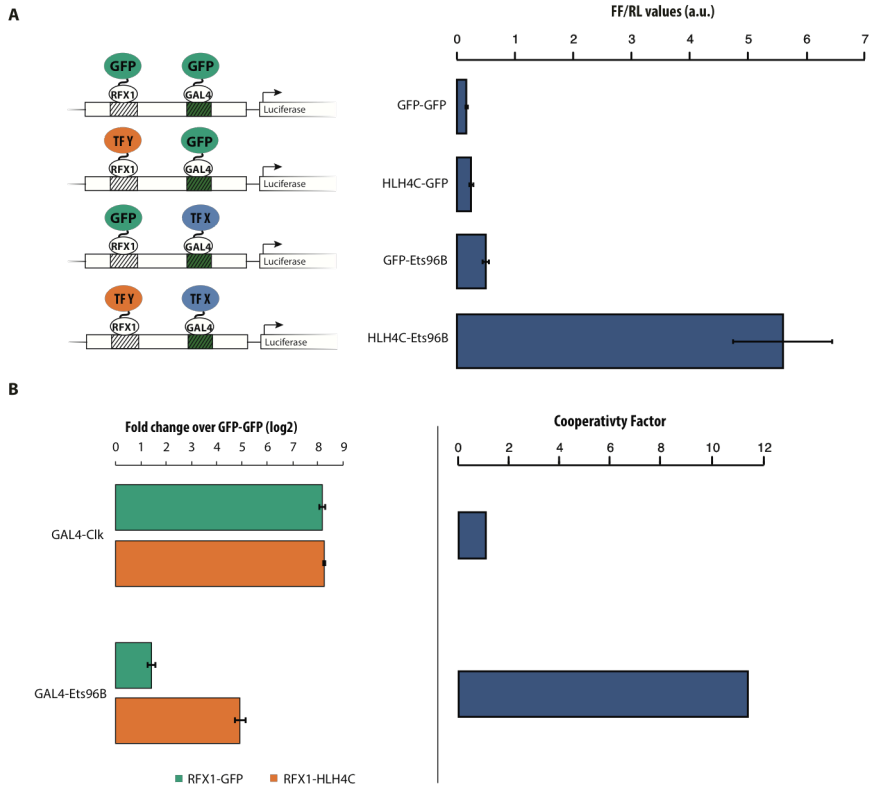

Fig 6: Cooperativity between TFs can be observed in a double recruitment assay, and be distinguished from non-cooperative activation. (a) The cooperative TFs HLH4C and Ets96B are not able to activate the reporter when either is co-recruited with GFP. Only when both are present at the reporter can we observe activation. (b) Global activators can be distinguished from cooperative activators by comparing their activity when recruited together with HLH4C or GFP. Global activators show a ratio of HLH4C-co-recruitment over GFP-co-recruitment (CF) of ~1 while cooperative activators have a CF value > 1.

Global activators are able to activate any context without the help of additional TFs. Therefore, it is important to distinguish between them and actual cooperative activation. A TF that relies on cooperativity will not be able to activate the reporter when co-recruited with GFP, whereas a global activator does not depend on a partner TF. This is not only true for the TF recruited to the EcR motif (here HLH4C), but also for the partner TF that we seek and which should function at the position of the GATA motif in combination with HLH4C but not alone. We therefore recruited all TFs via Gal4-DBD to the GATA motif position, once together with GFP at the EcR motif position and once together with HLH4C, and assessed transcriptional activation in both conditions. For each TF, the ratio of activation when co-recruited with HLH4C over co-recruitment with GFP (Cooperativity factor CF) specifically measures the cooperativity with HLH4C and is

around 1 for global activators and >1 for cooperative TFs – we used >1.5 and p<0.05 (t-test; n=4) as thresholds to identify cooperative TFs (Fig. 6b).

Cooperativity with HLH4C in the Nhe2 context

To test a c library of 474 Drosophila TFs for cooperativity with HLH4C, we used the Nhe2-EcR enhancer context and HLH4C as a bait. We found 40 TFs with a Cooperativity Factor (CF) larger than 1.5 and a significant p-value<0.05. The top two hits were Ets96B and Smox (Fig. 7), which are both from the same previously defined cluster as HLH4C (cluster 15). Looking more globally at all significant hits, we find TFs of the clusters 1,6 and 15 enriched among the 40 cooperative TFs (Table 1). 1 and 15 are the previously mentioned clusters of activators of hormone inducible enhancers. However, cluster 6 TFs exhibited a very different behavior in the single recruitment screen, in which they did not activate any of the 24 contexts and could therefore not be annotated as activating or repressing regulators. The new results from combinatorial recruitment with HLH4C now suggest that TFs from cluster 6 are obligate combinatorial and show cooperativity with HLH4C.

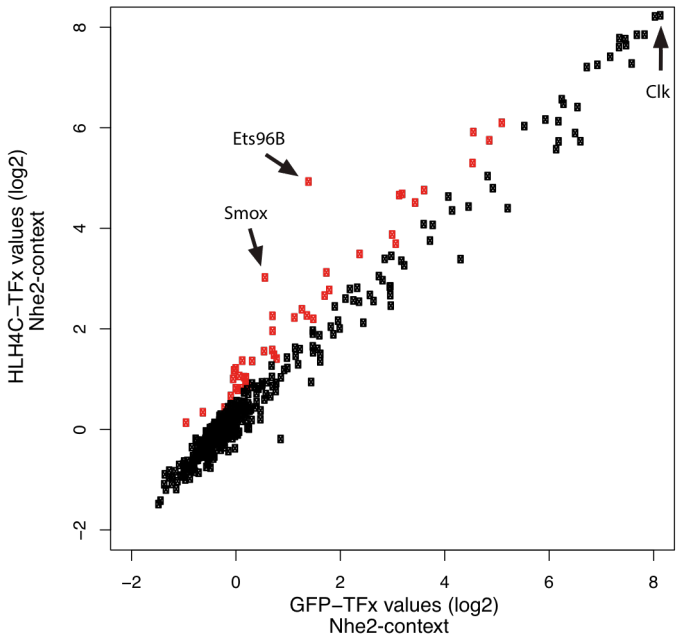

Fig 7: Testing 474 TFs for their cooperativity with HLH4C in the Nhe2-EcR-GATA context. 42 TFs showed cooperative activation when co-recruited with HLH4C (red; CF value >= 1.5, pvalue < 0.05). Indicated are the two top hits Ets96B and Smox (highest CF values, ~11 and ~5 respectively) and the global activator Clk (CF of 1).

Cooperative TF-pairs are identified independently of the bait TF and the specific enhancer context used

To confirm our results and test the dependency of the combinatorial TF pairs on the enhancer context and the bait TF, we changed these two parameters, one at a time. Changing the Nhe2 enhancer to an equivalent ecdysone responsive enhancer from the *singed* locus (sn) produced 272 hits (Fig. 8a). Of the TFs showing cooperativity in the Nhe2 enhancer, ~93% were also found in the sn context, which corresponds

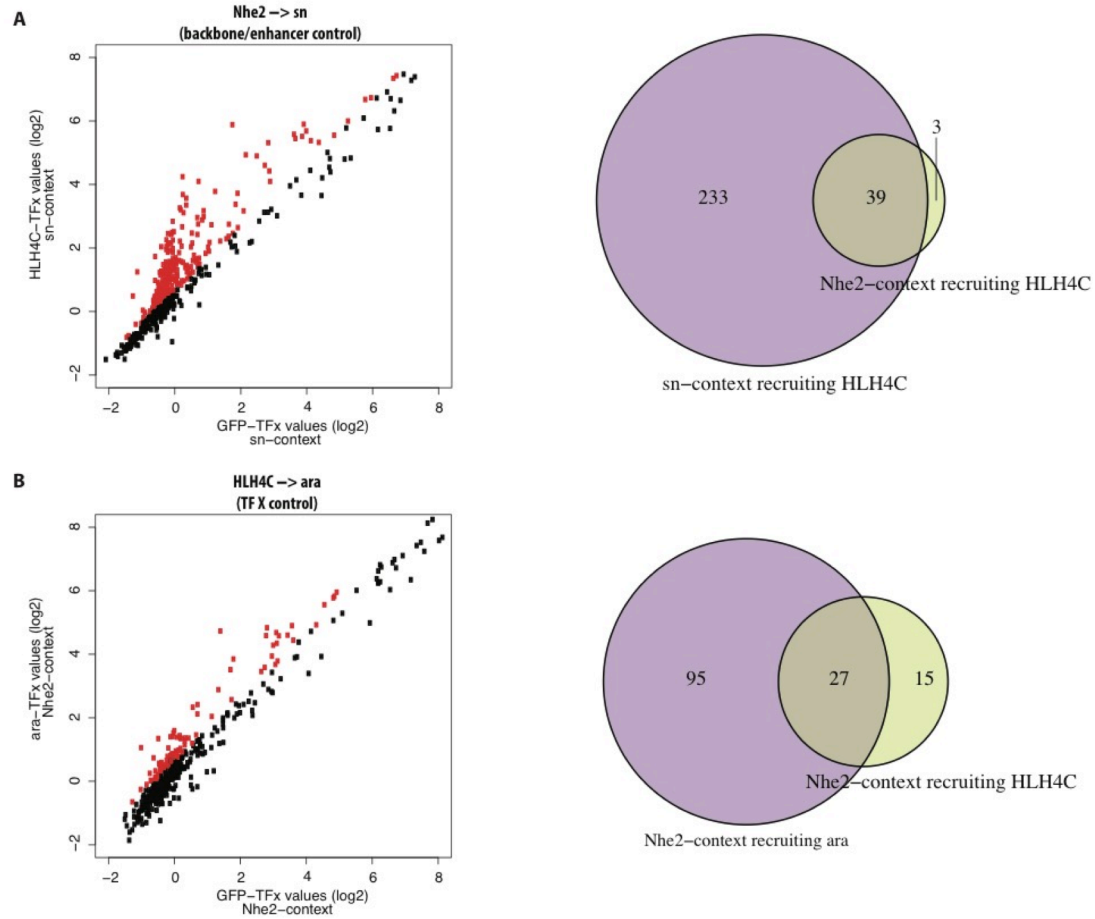

Fig 8: **Cooperatively activating TFs are consistently found when varying either the enhancer context or the bait TF.** (a) Changing the Nhe2 enhancer context to another ecdysone responsive enhancer (sn) finds 272 cooperatively acting TFs. 39 of the 42 TF found in the Nhe2 context are reproduced in this set up. (b) Changing the bait TF from HLH4C to ara finds 122 hits, 27 of which are also found in the HLH4C co-recruitment. The overlaps are highly significant (1.72-fold, p-value=$1.3 \times 10^{-08}$and 1.55-fold, p-value=5.6×10^{-09}, respectively).

to a highly significant enrichment (1.76-fold; p-value=1.3×10^{-08}) (Fig. 8a).

By changing the TF from HLH4C to ara, which is the TF with the most similar regulatory activity across all enhancer contexts according to the original single

recruitment screen (Stampfel et al., 2015), we identified 122 combinatorically acting TFs. These included 27 TFs that were among the 40 TFs found in the HLH4C co-recruitment (64%), again a highly significant enrichment (1.55-fold; p-value $5.6x10^{-09}$) (Fig. 8b).

The top hit from both the control screens was Ets96B as in the original screen (Smox was rank 3 in the sn-context screen and 8 in the screen with ara as a bait-TF). Overall, even though the original screen using HLH4C in the Nhe2-EcR context had fewer hits than both control screens, we could reproduce a significant fraction (93% and 64%) of the cooperatively acting TFs in both control screens (Fig. 8). To get a better understanding of which TFs act in this cooperative way, we identified which clusters (from the single recruitment screen, see (Stampfel et al., 2015)) these TFs belong to. Enrichment analysis for all three set ups found clusters 1, 6, 9 and 15 enriched in at least two of the three set ups (clusters 6 in all three) (Table 1).

HLH4C-Ets96B cooperativity is independent of the DBD used for recruitment

To exclude that the specific choice of DBD influenced the results on TF cooperativity, we swapped the GAL4 DBD for the LexA DBD and tested combinatorial activation for selected TF pairs . Using RFX1 to recruit HLH4C to the EcR motif position, resulted in the same synergistic activation when co-recruiting Ets96B with LexA rather than GAL4, showing that the cooperative effect is not dependent on the DBD GAL4 (Fig. 9). This suggests that the cooperativity between these TFs is not a consequence of the DBDs we use to recruit them.

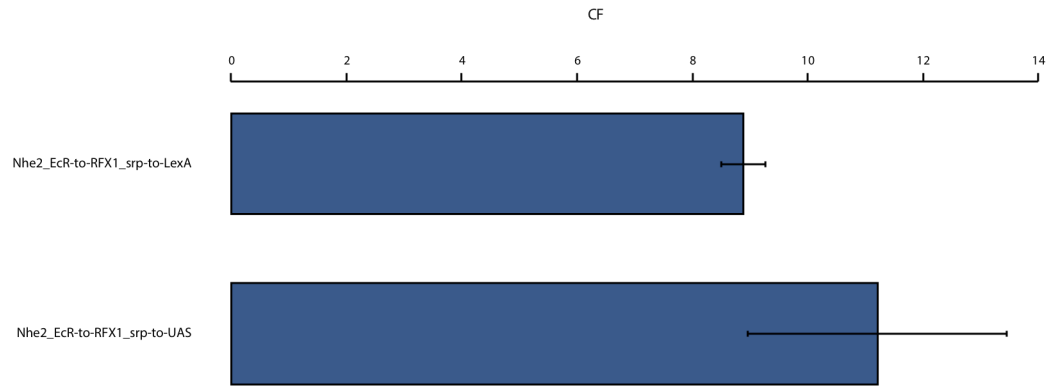

Fig 9: Cooperativity between TFs is independent of the DBD. Changing the DBD recruiting Ets96B at the GATA motif from GAL4 to LexA does not significantly influence (pvalue=0.051) its cooperative activation with HLH4C.

Testing the sufficiency of co-recruiting TF pairs in a synthetic context

The full enhancer context does not only consist of the two TFs we recruit using our two DBDs, but additionally of all TFs bound to the enhancer, the identity of which is unknown. To test if the TF pairs we found in the Nhe2-EcR-GATA context are sufficient for activation, we designed a synthetic reporter containing only two RFX1 and two UAS binding sites, thereby essentially removing all endogenously bound

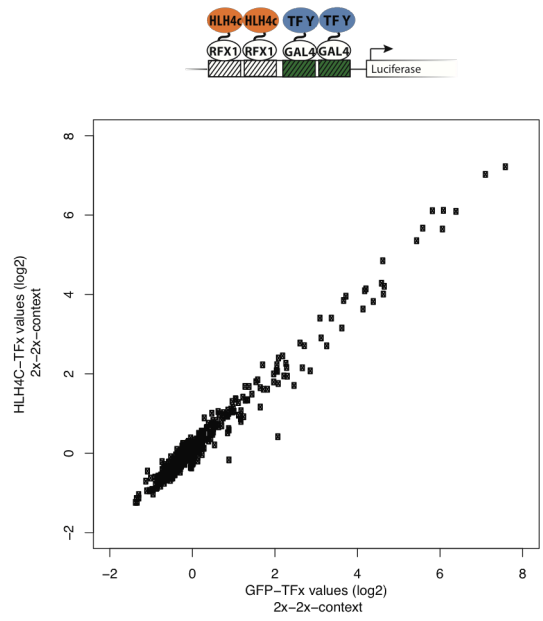

Fig 10: TF pairs are not sufficient for activation in a synthetic context (motif array). Co-recruiting TFs to a synthetic binding site array of 2 RFX1 binding sites followed by 2 UAS sites produces no cooperatively acting hits.

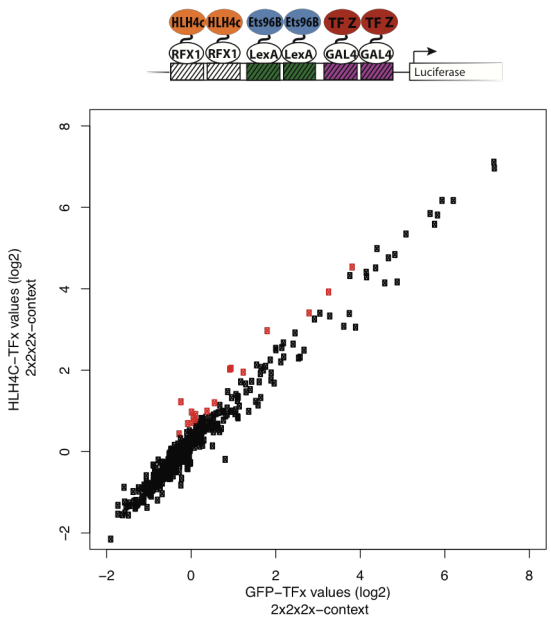

Fig 11: Triple recruitment of HLH4C-Ets96B and a third TF finds weak cooperativity in a synthetic context. Using the most cooperative pair found in the double recruitment in the Nhe2 context, HLH4C and Ets96B, as bait TF-pair and co-recruiting them with the 474 TFs, found 16 weakly cooperative TFs.

TFs. This allowed us to test if the cooperative activation we observe in the full enhancer context is driven solely by the two recruited TFs, or if it relies on additional factors. We could not find a single pair that is sufficient for activation (Fig. 10), suggesting either that more than two TFs are required or that the native arrangement of binding sites is important for TF cooperativity.

Testing the sufficiency of co-recruiting three TFs in a synthetic context

To address the possibility that a third TF is missing, we designed a synthetic reporter containing binding sites for RFX1, GAL4 and LexA, which enabled us to

simultaneously recruit three different TFs. From the double recruitment in the enhancer context we already knew that HLH4C-Ets96B is the best pair, that however seemingly requires additional TFs of the enhancer to function. We used this pair as a constant bait to look for a potential third partner. This resulted in weak cooperativity of 16 TFs with HLH4C and Ets96B. The highest CF (2.7, pvalue =$6.47x10^{-6}$) was observed with Dsp1 (Fig. 11).

Two TFs are not sufficient for combinatorial activation – even if their arrangement preserved

The arrangement of TF binding sites (i.e. order and spacing) has been shown to influence the strength and specificity of a reporter (Farley et al., 2015; Fiore and Cohen, 2016). To test if in our case loss of cooperativity resulted from the artificial motif arrangement in the synthetic contexts above, we transferred the endogenous arrangement into control DNA-backbones, preserving the order and spacing between the binding sites but changing the DNA sequence between them. We used a sequence from the Drosophila genome with GC content matched to enhancers, but with no enhancer activity in reporter assay, referred to as dmRan36. Additionally, we used a human enhancer and a Drosophila enhancer with no activity in S2 cells. Results for the different backbones were not uniform. Similar to the synthetic context above, the backbone from the background genomic sequence (dmRan36) did not show induction of the reporter (Fig. 12 a), neither for the co-recruitment of HLH4C-Ets96B nor HLH4C-ara. The human SMOC1 backbone exhibited the same behavior (no activation) for the HLH4C-ara co-recruitment, whereas the HLH4C-Ets96B pair was able to activate this context to nearly the same level as the original Nhe2 context. This discrepancy suggests two things: arrangement cannot be the sole determinant for cooperativity, otherwise the dmRan36 context should also be activated, and cooperativity of different TF pairs is influenced in degrees by the surrounding sequence. The third control backbone we used was a Drosophila enhancer sequence. This context could be activated by both co-recruitment pairs, suggesting it contains additional sequence features that facilitates activation. This variability between different backbones suggests that recruiting two TFs is not sufficient for activation, even if the motif arrangement and spacing is preserved.

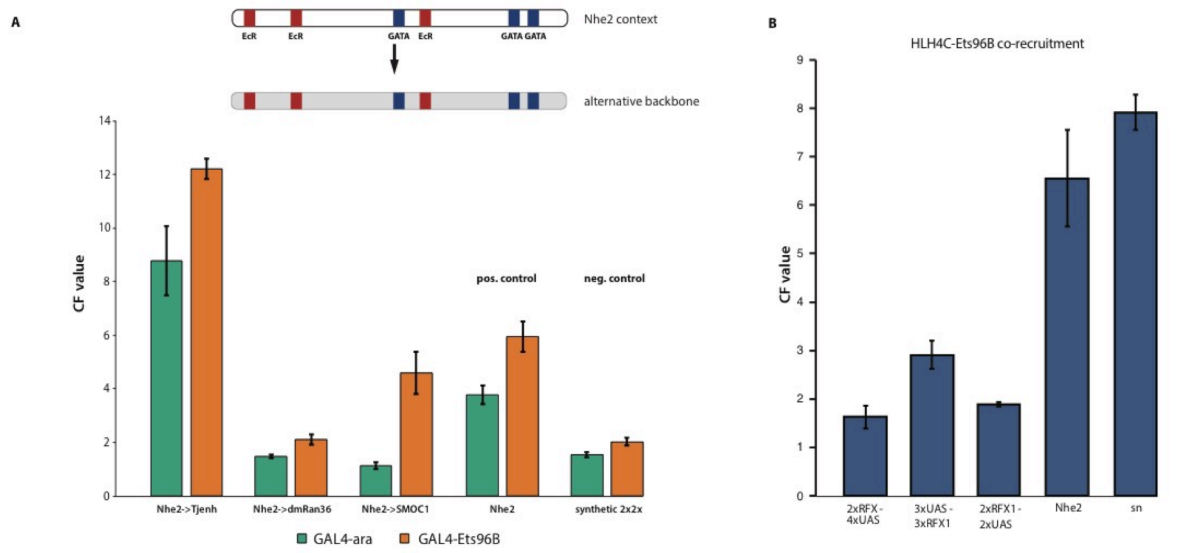

Fig 12: Testing the influence of motif arrangement on cooperativity. (a) We tested the influence that the motif arrangement has on cooperativity by preserving the number and relative distance between motifs but changing the sequence between the motifs (backbone). Different backbones had a different influence on the cooperativity. (b) Changing only the binding site number in a synthetic context does not have a strong effect on cooperativity.

In addition, to the motif arrangement test, we also tested how changing the DBD binding site number in the synthetic context influences cooperativity between HLH4C and Ets96B. We changed the synthetic context to contain the same number of binding sites (2 RFX1 plus 4 UAS and 3 RFX1 plus 3 UAS respectively) as the two full enhancer contexts (Nhe2 and sn) that we used previously. Although increasing the number of binding sites from 2 each to 3 each increased the signal, it did not reconstitute the activation seen in the Nhe2-EcR-GATA context (Fig 12b), further reinforcing the argument that number and arrangement of motifs alone is not sufficient for cooperativity.

The GAGA motif as a potential third contributor to combinatorial activity

Given the results above that motif arrangement cannot be the one determining variable for enhancer activity, we concluded that the three backbones must differ in ways that influences TF cooperativity and enhancer activity.

We therefore compared the motif content of the different backbones. The human SMOC1 enhancer contained 4 Fkh-like motifs, possibly explaining the more than endogenous activation of the reporter. However, this could not explain the difference between the dmRan36 and the endogenous Nhe2 enhancer. The only candidate with a good possibility of discriminating the active and inactive backbones was the GAGA motif, because there was a clear difference in number of GAGA motifs between the backbones that supported TF cooperativity and the one that did not (Table 2).

Co-recruiting 3 TFs in the Nhe2 enhancer

We designed a triple recruitment assay in the endogenous context, by mutating the GAGA motif additionally to the EcR and GATA motifs. Similar to the triple recruitment in the synthetic context we used HLH4C and Ets96B as the constant pair, and co-recruited 474 TFs in turn as a third TF. The screen revealed a total of 149 cooperatively acting TFs (Fig 13b). The TFs found in the Nhe2 context double recruitment overlapped to ~93% with the ones found in this experiment (3.3 fold-enrichment, pvalue = $3.6*10^{-20}$), however, there are certain differences. Clusters 8 and 14 are enriched in the triple but not in the double recruitment, whereas cluster 9 shows the opposite trend.

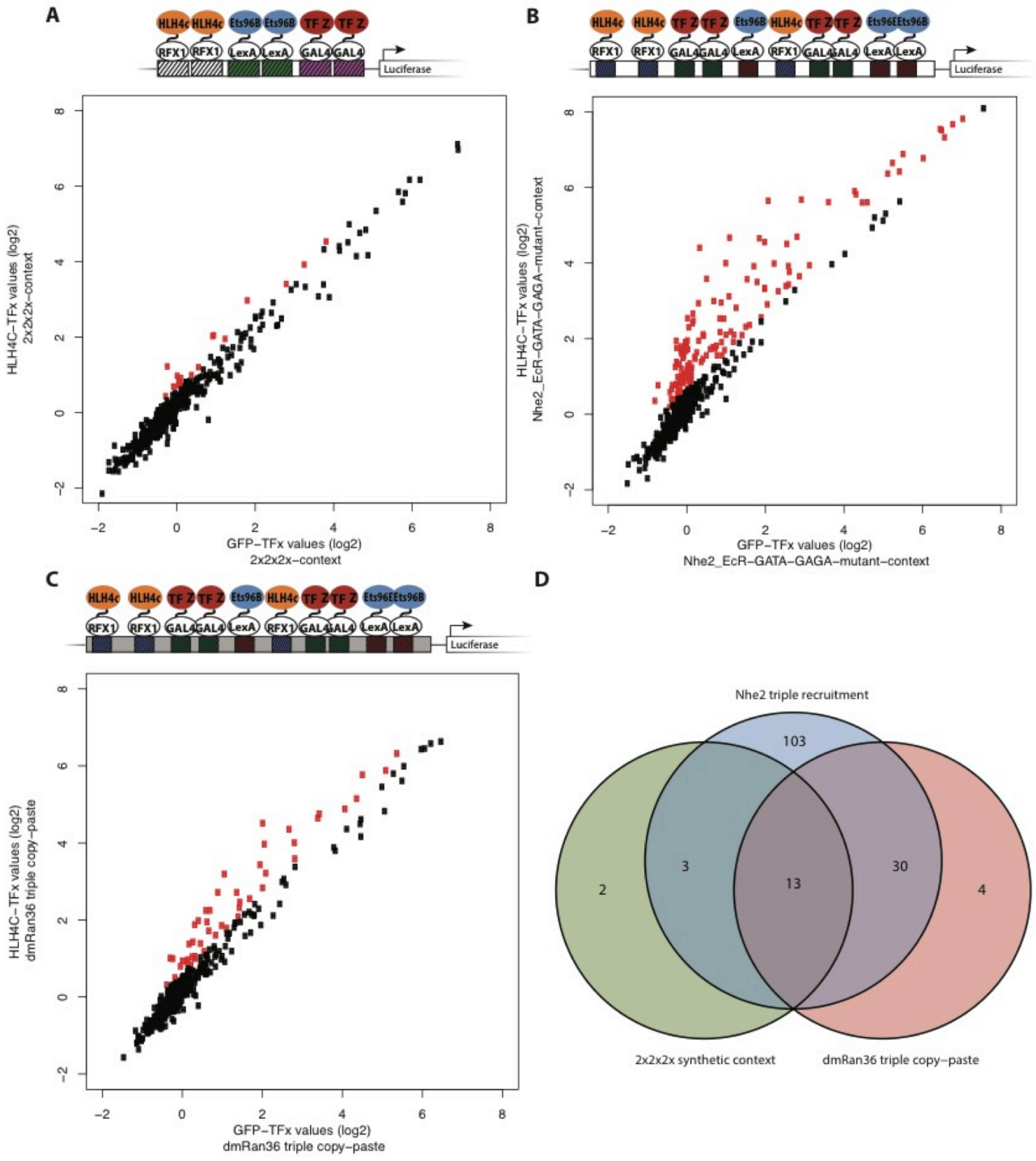

Fig 13: Triple recruitment of TFs produces reproducible hits in the full enhancer context as well as in the control backbone. (a) Triple recruitment in the synthetic context (same as Fig. 11, shown again here for comparison with (b) and (c)). (b) Triple recruitment in the Nhe2 context produces 149 cooperative TF hits. (c) The preservation of motif arrangement enhances the cooperativity seen in the synthetic context. We find 47 cooperative TFs. (d) Only 2 and 4 TFs respectively are found in the synthetic and copy-paste contexts that are not found in the full Nhe2 context.

More interestingly, when comparing the TFs found in the synthetic triple recruitment experiment with the endogenous enhancer context, we find 6 out of the 18 TFs are

also identified in the top 10 of the Nhe2-EcR-GATA-GAGA context (Fig 13d). Overall only 2 TFs that were found in the synthetic context were not identified in the Nhe2-EcR-GATA-GAGA context (~3.2-fold enrichment, pvalue = $7.5x10^{-8}$). This points to biologically relevant interactions being identified in the synthetic context. Interestingly Trl (GAGA) could be identified in the synthetic but only very weakly in the full Nhe2-EcR-GATA-GAGA enhancer context.

Since the qualitative cooperativity in the synthetic context seems to be present between the same factors as in the full enhancer but to a lower degree, we went back to testing if the arrangement of binding sites could amplify the signal. Even though binding site arrangement was not sufficient to promote cooperativity in the double recruitment experiment, it might have a quantitative effect that would let us detect stronger cooperativity between triplets. We therefore designed a reporter similar to the ones used for testing the influence of motif arrangement in the double recruitment before, and transferred the arrangement of the EcR, GATA and GAGA motifs of the Nhe2 enhancer into the dmRan36 backbone sequence. We then measured the cooperativity of all TFs recruited to the positions of the GAGA motifs with HLH4C and Ets96B, recruited to the positions of the EcR and GATA motifs, respectively. We found 47 cooperatively acting TF 43 of which overlapped with the ones found in the full Nhe2-EcR-GATA-GAGA context (Fig 13 c and d). The levels of activation were higher than the ones found in the completely synthetic context, but lower than in the full Nhe2 context, reinforcing the notion that arrangement promotes the cooperativity.

Discussion

In this work I present a multidimensional recruitment assay, that allows the assessment of specific interdependencies between transcription factors, and find pairs or triplets of TFs that can specifically work together to activate transcription. I developed a set up in which we could probe cooperativity of 474 TFs with one or two bait TFs. Previous experiments identified a group of TFs that act in an obligate combinatorial manner (Stampfel et al., 2015), i.e. they are only able to elicit their function if their correct partner TFs are present at the same enhancer, letting me probe which TFs they are able to work together with.

Recruitment assays have been widely used to study the role of TFs in transcription control (Ptashne and Gann, 1997). They have been successfully used to study the activity of single TFs (Bryant and Ptashne, 2003), and pairs of TFs (Cheng et al., 2004; Keung et al., 2014). These approaches have identified synergistic interactions between e.g. different chromatin remodelers. In my thesis I wanted to expand this approach, to investigate TFs that do not activate transcription when recruited on their own to a naïve (synthetic) context but can strongly do so in combination with other TFs.

I consistently observed the same cooperatively acting TFs when using equivalent enhancer contexts and bait TFs. Furthermore, was I able to demonstrate that these cooperative pairs are not sufficient for activation, but TF triplets can activate gene transcription at a low level. This low-level activation is enhanced when the arrangement of motifs from the enhancer context is preserved, and even further enhanced if the intervening sequence corresponds to an enhancer sequence. Overall this work demonstrates, that cooperative activation of TFs, even though it is facilitated by arrangement and number of the binding sites, relies on the specific identity of the partner TFs. These cooperative interactions lead to super-additive activation, and can therefore not be explained by a simple model in which every TF contributes a certain activating potential to the overall enhancer activity that corresponds to the sum of all individual contributions. Instead, these TFs each bring a distinct function (unknown to us), that is required for enhancer function.

Reasoning that TFs with different functions complement each other and would therefore cooperate at the enhancer, we initially expected to find TFs from different functional clusters (as defined by Stampfel et.al.,2015) to pair with HLH4C (cluster 15). This was however, not the case, as two of the most enriched cluster of TFs were 1 and 15, which behave very similar and are equivalent to HLH4C. It is of course possible that enhancers bind classes of TFs that are partly defined by their ability to work together. If the ability to activate in a certain enhancer context is coupled to the ability to cooperate with the other TFs bound to this enhancer, it would explain why they would even cluster together in a single recruitment assay.

One of the most intriguing findings in this aspect for me was, however, the presence of TFs from cluster 6. TFs from this cluster did not show activation of any of the original 24 contexts in the single recruitment assay. It is particularly interesting as this cluster is the biggest one (contains more TFs than any other) out of all 15 clusters. Since I was able to detect cooperative activation of some TFs of cluster 6 with HLH4C and ara, it suggests that these TFs have a stricter cooperative requirement than TFs from any other cluster, and that the cluster is not so much defined by its activity as more by the lack of correct partner TFs. The TFs from this cluster working together with the bait TFs used in this study would therefore be a subgroup of this cluster. Other TFs from this cluster could cooperate with different bait TFs, splitting this large cluster into many smaller ones. What they all probably have in common is their obligate cooperative behavior.

The other main finding is that cooperative TF pairs are not sufficient for activation when taken out of the enhancer context, even when adjusting the number of binding sites for each TF. Going a step further and also preserving the motif arrangement demonstrated that this too is not sufficient to promote cooperative activation between two TFs. Since we used a limited number of backbone sequences for these experiments, the influence of the sequences intervening the motifs must be considered. The first sequence we chose was tested to have no enhancer activity, but was matched in its GC content with other Drosophila enhancer sequences. Additionally, we used two enhancer sequences, one from human (SMOC1) and one from Drosophila that is however not active in the cell type we were using for our experiments (S2 cells). The reasoning for using these sequences, is that we wanted

the sequence composition, and its ability to e.g. be packaged into nucleosomes to be as close to that of a functional enhancer as possible. The fact that these sequences are functional enhancers (active in cells other than S2 cells), makes it very likely that additional TF binding sites are present in the sequences. By inserting the DBD binding sites into the sequence it is also possible that we created new TF binding sites that could influence the activity of the sequence. In addition to TF motifs, other factors might contribute to the differential behavior of the various backbones. For one, enhancer sequences might acquire and position nucleosomes differently than other genomic sequences (Radman-Livaja and Rando, 2010). Abundant nucleosomes might mask TF binding sites and therefore hinder the binding of the protein to the DNA. Another reason for the observed differences might be the DNA shape. It has been shown that protein-DNA interactions not only rely on the specific underlying sequence, but also the local and global shape of the DNA (Rohs et al., 2010). However, since these parameters are much harder to test in our set up, we did not include them in our experiments but instead decided to focus on motifs for additional TFs that might be differentially present in the different backbones. The ideal way to set up this experiment would be to clone a large number of different intervening sequences and test them in a pool to avoid any bias that might stem from any specific sequence. However, since our enhancer context is approximately 500bp long, it is not easy to do this. Despite these shortcomings, our results show that motif arrangement alone is not the determining factor for cooperativity between TFs. It does, however, seem to have a quantitative effect. In the triple recruitment experiments I found a larger number of cooperative TFs when preserving the motif arrangement. The CF values for the individual pairs were overall also higher than in the purely synthetic context. Likely there is a hierarchy to TF cooperativity. Cooperative activation first and foremost depends on the identity of the co-recruited TFs, and can be enhanced be providing a more ideal motif arrangement. Providing additionally intervening sequences containing motifs recruiting other cooperative TFs further enhancer the activation. These more subtle changes, like motif arrangement might be important fine tune the expression of genes in vivo (Farley et al., 2015).

Cooperativity between TFs can happen at different levels. Cooperativity at the level of binding for example, can arise from TFs that change their binding preference depending on whether they bind together with a partner TF or alone (Jolma et al.,

2015). Post binding cooperativity is less well understood. The second class of proteins involved in transcriptional control, transcriptional cofactors (COFs) are likely involved.

TFs exert part of their function by recruiting COFs. It is possible that combinatorically active TFs are not able to recruit the necessary COFs alone but only in combination with their partner TFs. It would be interesting to test this hypothesis, by looking at the whole collection of proteins that assemble at given enhancer sequence. The expectation for a model like described above, would be that we after mutating the binding site for a cooperative TF we would not only lose binding of this particular TF but also, depending on the mode of cooperativity, either binding of its partner TF directly or their recruited COFs. A Mass Spectrometry approach would make such a study possible. By doing a pull-down experiment on different mutant versions of an enhancer sequence, we could learn more about how TFs assemble at the sequence. Expanding this idea, it would also be possible to look at the communication between enhancers and core-promoters (CPs) by providing both on the DNA template. We already know that enhancers and CPs are divided in two big functional classes specific to either developmental or housekeeping genes (Zabidi et al., 2015). It would be interesting to see if different TF combinations mediate the communication between this two types of enhancers and CPs, and if we can subdivide these classes by looking at the proteome that they bind.

In conclusion, we established a co-recruitment assay for up to three TFs that can in theory be extended to observe the cooperativity between any number of TFs. We observed cooperativity between up to three TFs, revealing that this number of TFs is able to active transcription of a reporter even when recruited outside of an enhancer context, i.e. these triplets are sufficient for activation, and reconstitute a cooperatively acting enhancer.

Enhancer biology is an important and interesting field of study, not least because changes in enhancer sequences can lead to developmental defects like polydactyly. Since TFs are the proteins that read the enhancer-encoded information, we have to understand which TFs bind to which enhancers and how they subsequently elicit their activating or repressing functions. One aspect of this is TF cooperativity, both at the level of TF binding, and after binding. Understanding which TFs are able to work

together would enable us to design enhancers with specific expression patterns. Overall it is important to study enhancers from both the sequence level and the proteins that bind to that sequence to get a comprehensive understanding.

Tables

	Nhe2_EcR-toHLH4C		sn_EcR-to-HLH4C		Nhe2_EcR-to-ara	
Cluster	**Enrichment**	**p-value**	**Enrichment**	**p-value**	**Enrichment**	**p-value**
15	5.31	3.29x10-9	1.43	3.34x10-3	1.53	0.066
1	4.72	8.63x10-5	0.76	0.934	3.31	1.95x10-7
6	2.48	0.004	1.77	1.16x10-12	1.91	4.45x10-4
9	0.34	0.957	1.67	2.01x10-7	1.69	0.020
13	0.75	0.768	1.37	0.011	1.45	0.111
14	0.65	0.804	0.95	0.697	0.23	0.994
12	0.39	0.975	1.31	0.002	0.81	0.849
8	1.08	0.570	0.59	0.992	0.38	0.983
2	0.00	1.000	0.76	0.934	0.41	0.973
3	0.00	1.000	0.20	1.000	0.00	1.000
4	0.00	1.000	1.29	0.010	1.07	0.443
5	0.00	1.000	0.00	1.000	0.36	0.959
7	0.00	1.000	0.30	1.000	0.63	0.955
10	0.00	1.000	0.25	1.000	0.46	0.975
11	0.00	1.000	0.53	1.000	0.17	1.000

Table 1: Cooperatively acting transcription factors from the three double recruitment screens are enriched in clusters 15, 1, 6 and 9. Clusters were previously defined, see Stampfel et.al.. The cluster for global activators (Cluster 8) is not significantly enriched in either of the three screens.

TJ enhancer			**SMOC1**		
motif	**times**	**motif sequence**	**motif**	**times**	**motif sequence**
srp / GATA	2	GATA	srp / GATA	4	GATA
Trl / GAGA	6	GAGA	Trl / GAGA	3	GAGA
CACA	5	CACA	CACA	6	CACA
Tj	4	GCTGA	Tj	1	GCTGA
Fkh-like	4	TAAACA	Fkh-like	1	TAAACA
dmRan36					
motif	**times**	**motif sequence**			
srp / GATA	1	GATA			
Trl / GAGA	2	GAGA			
CACA	8	CACA			
Tj	2	GCTGA			
Nhe2			**sn**		
motif	**times**	**motif sequence**	**motif**	**times**	**motif sequence**
srp / GATA	3	GATA	srp / GATA	4	GATA
Trl / GAGA	4	GAGA	Trl / GAGA	3	GAGA
CACA	3	CACA	CACA	5	CACA
Tj	1	GCTGA	Tj	3	GCTGA
			GC	1	GCGC

Table 2: The backbone DNA sequences used to preserve the motif arrangement but vary the intervening sequences contain TF binding sites.

Materials and Methods

Drosophila S2 cell transfection.

Drosophila S2 cells were transfected using jetPEI (peqlab 13-101-40N). Cells were seeded 4 hours before transfection in clear polystyrene 384-well plates (ThermoScientific 164688) at a density of 30.000 cells per well (30µL of a $1*10^6$ cells per mL suspension). Each well was transfected with 36ng DNA. For double recruitment 27ng of the firefly reporter, 3ng of Renilla expressing plasmid and 3ng of each DBD-TF fusion construct was used in a total of 7.5µL EB buffer. For triple recruitment reporter concentration was adjusted to 24ng, while Renilla and DBD-TF fusion constructs were kept at 3ng totaling 10.35µL. The DNA mix was filled up to 15µL with 150 mM NaCl (polyplus) and prepared in 96-well plates. 15 µL of the transfection reagent were added to each well (13.95 µl 150 mM NaCl, 1.05 µl jetPEI) and mixed. After 30min incubation at room temperature, the cells were transfected in quadruplicates. 6µL of the transfection mix was added to 4 adjacent wells of a 384-well plate for each construct. Luciferase assays were performed 48 hours after transfection. Pipetting of the transfection reagent onto the DNA mix and all subsequent steps were carried out by the Bravo Automated Liquid Handling Platform (Agilent).

Cloning of destination vectors pAGW-RFX1-DBD.

The RFX1-DBD was amplified from genomic yeast DNA using the primer 5'-CCAGATATCATGCCACCCAAGAAGAAGCGAAAAGTAGAAGATCCATCCATTTCAAGGAGAAACACGCAAGAAATAATTG-3' containing a nuclear localization signal and 5'-ATACCGGTGGCCGCCGCCAGGTACCTGCGGCGAAGGTGAAG-3'. Amplification was performed using the Q5 High-Fidelity DNA Polymerase (NEB). The PCR product was digested with the restriction enzymes AgeI and EcoRV (both NEB) and ligated into the pAGW destination vector which was digested using the same enzymes.

Cloning of luciferase reporter vectors.

Gateway compatible reporter plasmids were taken from Stampfel et al. Synthetic binding site arrays for GAL4, RFX1 and LexA were generated using self-annealing primers containing the attB sites for gateway cloning. Motif mutant enhancer contexts were ordered as synthesized fragments (IDT). For the identification of positive clones Sanger Sequencing with primers 5'-TGTGAATCGATAGTACTAACATACG-3' and 5'-CAACTGATGCTCTCAGCCACCCCG-3' was used.

N-terminal GAL4-DBD-tagged TF library.

All GAL4 tagged TFs were taken from previously prepared stocks, see Stampfel et al. 2015.

N-terminal RFX1-DBD-tagged LexA-DBD-tagged TFs.

Gateway compatible entry clones containing the open reading frame without the stop codon for each TF were taken from Stampfel et al. Act5C-promoter driven expression clones were generated using the gateway system. TFs were shuttled into the DBD containing destination vectors (pAGW-RFX1-DBD and pAGW-LexA-DBD) by an LR reaction. LR reactions were set up by mixing 150ng of the DBD-containing destination vector with 150ng of the TF-containing entry clone and 1µL LR clonase II enzyme mix (Invitrogen). Positive clones were identified by Sanger Sequencing using primers 5'-GGATACTCCTCCCGACACAA-3' and 5'-CACACCACAGAAGTAAGGTTCC-3'.

Luciferase Assays.

Luciferase assays were performed as previously described (Stampfel et al. 2015), using self-prepared D-Luciferin (GoldBio LUCK-250) and Coelenterazine (pjk-GMBH 102111) substrates and lysis buffer (as described in Hampf et al. 2006). For cell lysis, supernatant was first removed, 30µL of lysis buffer was added to each well and

incubated at room temperature while gently shaking. After lysis 10µL of the lysate were transferred to a black 384-well plate for luminescence assays (Nunc MaxiSorp, Sigma-Aldrich P6491-1CS). Pipetting steps were performed by the Bravo Automated Liquid Handling Platform (Agilent). Luminescence was measured for Firefly and Renilla after adding 20µL of each substrate using a Biotek Synergy H1 plate reader in combination with a plate stacker.

Luciferase data analysis.

All Firefly luciferase signals were normalized to Renilla signals to control for transfection efficiency and cell number. Further all TF co-recruitment values were normalized to co-recruitment of GFP-GFP (Fold change over GFP). Cooperativity between TFs was assessed by comparing their co-recruitment with GFP to their co-recruitment with HLH4C. Transcription factors were called “cooperative” if the ratio of HLH4C-TF co-recruitment with GFP-TF co-recruitment was > 1.5.

References

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Bryant, G.O., and Ptashne, M. (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. Molecular Cell *11*, 1301–1309.

Cheng, J.X., Gandolfi, M., and Ptashne, M. (2004). Activation of the Gal1 Gene of Yeast by Pairs of “Non-Classical” Activators. Current Biology *14*, 1675–1679.

Crocker, J., Ilsley, G.R., and Stern, D.L. (2016). Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nat. Genet. *48*, 292–298.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328.

Fiore, C., and Cohen, B.A. (2016). Interactions between pluripotency factors specify cis-regulation in embryonic stem cells. Genome Res. *26*, 778–786.

Gertz, J., Siggia, E.D., and Cohen, B.A. (2008). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature *457*, 215–219.

Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jiménez, F., Baylies, M.K., and Michelson, A.M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell *103*, 63–74.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell *38*, 576–589.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388.

Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., and Khalil, A.S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell *158*, 110–120.

Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. Current Biology *20*, R754–R763.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Reiter, F., Stampfel, G., Stark, A. (2016). Poster for the Conference "Cell Symposia Transcriptional Regulation in Development and Disease", June 26-28, 2016 Chicago. (The Abstract printed here was taken from a poster submitted by me (Franziska Reiter) to the above mentioned conference)

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. *43*, 73–81.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. *79*, 233–269.

Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryń, L.M., Arnold, C.D., and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Molecular Cell *54*, 180–192.

Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45*, 1021–1028.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proceedings of the National Academy of Sciences *86*, 5434–5438.

Yáñez-Cuna, J.O., Arnold, C.D., Stampfel, G., Boryń, L.M., Gerlach, D., Rath, M., and Stark, A. (2014). Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. Genome Res. *24*, 1147–1156.

Yáñez-Cuna, J.O., Kvon, E.Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. Trends Genet. *29*, 11–22.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559.