Identification and validation of maize enhancers

A cartography of the maize regulatory genome

Weber, B.B.M.

General Discussion

A bird’s eye view on the maize regulatory genome

General Discussion

A bird's eye view on the maize regulatory genome

The work described in this thesis is dedicated to the identification and characterization of plant enhancers. Enhancers can be defined as transcriptional regulation units. They are DNA sequences that serve as docking platforms for transcription factors that influence the expression of target genes, and thereby allow the establishment of particular expression patterns in time and space. Plant enhancers remain astonishingly poorly described in comparison to their animal counterparts for which numerous genome-wide analyses have revealed their associated proteins, chromatin and DNA characteristics. Through the identification and characterisation of enhancer sequences in *Zea mays*, this work, aimed to reduce the gap in knowledge on plant enhancers. Below, we discuss the main results presented in the different chapters.

Plant and animal enhancers: similar but different

The first chapter of this thesis summarizes the current knowledge on plant and animal enhancer characteristics and provides a list of previously identified enhancers in the plant kingdom [1]. Examination of characteristics associated with plant and animal enhancers reveals many similarities between enhancers from the two kingdoms. At active enhancers, these include high chromatin accessibility levels, low DNA methylation, and enrichment for specific histone marks such as H3K9ac and H3K27ac [1,2].

There are, however, also differences between plant and animal enhancers. For example, in animals, preferential enrichment of H3K4me1 over H3K4me3 has been described as a mark to distinguish enhancer sequences from transcription start sites (TSS), independently of their activity status [3–5]. Notably, we show that in maize tested candidate enhancers are mostly not enriched for H3K4me1 [6].

The histone marks H3K9ac and H3K27ac are hallmarks of active enhancers in animals [4,7,8]. In rice, relatively high levels of H3K9ac and H3K27ac were observed at intergenic regions, possibly indicating the presence of active enhancers [2,9]. In maize, H3K9ac as well as H3K27ac enrichment at the *b1* enhancer correlated with the activity of the enhancer [6]. Still, H3K9ac appeared to better distinguish the inactive from the active state of the *b1* enhancer. In *Arabidopsis*, the active candidate enhancer sequences were poorly associated with H3K27ac [10], suggesting differences between plants and animals regarding histone marks associated with active enhancers. In their inactive state, the same candidate *Arabidopsis* enhancers were

associated with H3K27me3 [2,10]. In animals, H3K27me3 is replaced by H3K27ac upon enhancer activation [7,11,12], and it is therefore tempting to think that the same occurs in plants. The poor association of active enhancers with H3K27ac in *Arabidopsis* may be the result of an experimental artefact.

In animals, active enhancers have been shown to be characterized by the production of low levels of short, and relatively unstable transcripts coined enhancer RNAs (eRNAs) [13–15]. In addition, eRNA producing-enhancers were shown to be transcribed in a bi-directional fashion. In plants, a previous study on nascent transcripts in both *Arabidopsis* and maize suggested the absence of transcription at enhancers [16]. We found, by contrast, that the candidate enhancers identified in Chapter 2, were transcribed (see chapter 4). Genome-wide, the candidate enhancer-associated transcripts were produced in a uni-directional fashion; only very few bi-directionally produced transcripts were detected from candidate enhancers ([6], unpublished CAGE-seq data). These results are, to our knowledge, the first evidence for the existence of plant eRNAs, but also highlight the differences between plant and animal enhancers.

Predicting enhancer sequences

In the opening chapter of this thesis, we review some of the features associated with enhancer sequences and techniques allowing their identification. It is important to realize that individual techniques do not allow for the identification of enhancers *per se* but instead allow to "map" the distribution of specific feature(s) associated with enhancers. Techniques such as DNaseI-seq, MNase-seq or ATAC-seq can map accessible regions for instance, while ChIP-seq can map specific histone modifications and the binding of transcription factors, and Bisulfite-seq low- and unmethylated regions (see chapter 1 for more techniques and details). Also, each technique has an intrinsic predictive value, which depends on the degree of association of the studied characteristic with enhancers. For example, accessible chromatin is found at both poised and active enhancers, but also at promoters [11,17–20]; histone marks associated with active enhancers (*e.g.* H3K27ac or H3K9ac; [4,8]) or inactive enhancers (*e.g.* H3K27me3; [7,11,12]) can also mark active and inactive promoters, respectively [4,9,21,22]. In addition, for the same activity status, distinct enhancers might exhibit different histone marks [23]. Importantly, there is no single chromatin characteristic that is sufficient on its own to predict enhancers, reveal their activity status and distinguish them from other genomic sequences. Genome-wide identification of enhancer sequences would therefore benefit from combining multiple chromatin features. The resulting maps, like individual layers, can then be overlapped to define candidate enhancer regions. Note that maps remain an approximation of the reality or, as coined

by Alfred Korzybski "the map is not the territory". With this statement, Korzybski wished to illustrate the difference between reality (the territory) and belief (the map, the representation one has of a given reality). In a similar way, genome-wide maps of features and the resulting candidate enhancer regions we define based on the maps remain approximations of the exact enhancer "territory", an estimate of the enhancer's position in the genome.

Genome-wide identification of enhancers ideally takes into account as many tissues and developmental stages as possible, as this would enable the identification of tissue-specific enhancers. However, analysing numerous cell types and developmental stages, and ideally also in various environmental conditions is very expensive. Instead, combining existing experimental data with *in silico* methods to uncover regulatory sequences is appealing. Classically, such approaches are based on the identification of sequence conservation or presence of Transcription Factor Binding Motifs (TFBM) [24–27]. Hereby, the results are very much dependent on the evolutionary distance of the species compared, the quality of consensus TFBMs and the ability of transcription factors (TFs) to bind such motifs [28–30]. The last decade has seen rapid breakthroughs in machine learning approaches, especially with image analysis for which extremely large data sets are available (for potential applications in the medical field see [31,32]). Machine learning-based methods were also used to predict enhancer sequences [33–38]. Only, a *sine qua non* condition to the success of such approaches is the collection of numerous datasets and good reference genomes. So far, those are still relying on human contributions! In plants, only a few reports using machine learning-based prediction of *cis*-regulatory elements are currently published [39–41]. Prediction of enhancer sequences using machine learning will certainly become a major research area in the near future and is of interest to define the best set of features for such prediction.

Predicting candidate enhancers in *Zea mays*

In Weber *et al* ([1]; Chapter 1), we call for large-scale discovery of plant enhancers. To match words with actions, we focused on the genome-wide identification of candidate enhancer sequences in the genome of the B73 reference line ([6]; Chapter 2). This was done by looking for genomic regions with high chromatin accessibility (DNaseI-seq), low levels of DNA methylation (BS-seq) and enriched levels of H3K9ac (ChIP-seq), as these characteristics were shown to be associated with active enhancers in both plants and animals [1,2]. The study performed in two different tissues (V2-IST and husk leaves) identified 1,702 intergenic regions defined as candidate enhancers ([6] and unpublished results). More specifically, 472 and 1,500 candidates were identified in V2-IST and husk, respectively; 270 candidates were predicted to

be active in both tissues. Importantly, the method successfully identified three (out of five) validated or putative enhancers (enhancer of *b1* and *tb1*, and DICE, the putative enhancer of *bx1*). Note that the putative enhancer *Vgt1* was not part of the predicted enhancer sequences (see below and Chapter 3 for details).

1,702 candidate enhancers is a relatively small number of identified candidates compared to the numbers found in other studies such as in *Arabidopsis* (about 10,000 enhancers), or humans (about 40,000 enhancers) [10,14]. Several factors can explain this discrepancy. In *Arabidopsis*, the prediction of intergenic candidate enhancers was done in two tissues, but solely based on chromatin accessibility [10]. In our study, a DNase I Hypersensitive Sites (DHS)-based prediction would result in the identification of approximately 9,200 candidate enhancers across both investigated tissues, a similar number as found in *Arabidopsis*. Recently, a study in tomato identified about 35,000 gene-distant DHSs across two different developmental stages [42]. This number drops to about 1,000 candidate regulatory regions across the two developmental stages when only sequences are considered that are marked by both a DHS and H3K4me3, which is equivalent to the number of candidate enhancers found in our study. Note that H3K4me3 is a mark that is shown to be more associated with a gene's TSS than with enhancer sequences [3–5]. In the absence of details on the level of H3K4me3 enrichment at these gene-distant DHSs, it is unclear whether they represent candidate enhancer regions, alternative promoters or unannotated TSSs of coding genes. The higher numbers of identified enhancers in human can for a large part be explained by the use of 120 different cell types; our analysis was based on only two different tissues. In addition, in our study, BS-seq identified 43,000 intergenic DNA regions as low or un-methylated (LUMRs). In plants, levels of DNA methylation appear to be relatively stable across tissues and stages ([43,44]; Oka, Stam and Springer unpublished results). Thus, we hypothesized that regions defined as LUMRs in our study could therefore represent the complete pool of intergenic *cis*-regulatory elements in B73. Data described in Chapter 4 suggest this may indeed be the case.

Strikingly, candidate enhancers identified in Chapter 2 were generally characterized by an asymmetric enrichment of H3K9ac. This result contrasts with previous reports in animals, indicating a symmetric distribution of H3K27ac at active enhancers (*e.g.*[7,14]). In animals, this symmetric distribution is correlated with the bi-directional production of eRNAs at the enhancer site [13,14]. We hypothesize that the asymmetric distribution of H3K9ac at enhancer candidates is connected to the relatively high and low number of uni- and bi-directional eRNAs observed in our maize tissues, respectively ([6], unpublished results). The asymmetric

enrichment of H3K9ac and eRNAs discriminates the regulatory sequences from maize and animals.

With an asymmetric enrichment of H3K9ac, high levels of accessibility and low levels of DNA methylation, the chromatin profile of candidate enhancers resembles the one of transcriptional start sites of genes. However, compared to TSSs, candidate enhancers are generally characterized by lower levels of transcripts and H3K4me3, and slightly higher levels of CG and CHG methylation at the sequences flanking the candidates ([6], Chapter 2). These observed similarities in the chromatin profiles of promoters and enhancers reinforce the idea that these two types of sequences are two sides of the same coin: transcriptional units [45]. In addition silencers and insulators seem to share similar characteristics with enhancers and promoters [46–51]. This was also indicated in our study in which candidate enhancer sequences appeared able to act as tissue-specific enhancers, but also as insulators or transcriptional silencers (Chapter 4). Sometimes these different functions appeared combined in one and the same sequence. The latter results are supported by publications highlighting the fact that *cis*-regulatory sequences can have multiple functions [45–47] and therefore questions the idea of strict distinctions between definitions of promoter, enhancer, silencer and insulator. For this reason, the candidate enhancers from chapter 2 will below be defined as candidate regulatory sequences.

The boundaries of the candidate regulatory sequences identified in chapter 2 were defined as the limits of the LUMRs regions. We mentioned that a map is an approximation of the "territory". Our candidate regulatory sequences are an estimate of the exact regulatory sequence position; exact coordinates remain to be identified. Do they include DNA sequences that are not absolutely required for the regulatory activity? To answer such question, one would need to dissect candidates into segments and test the regulatory abilities of each segment (and combination of segments) in as many tissue/cell/developmental conditions as possible. This could be achieved using methods such as massively parallel reporter assay (MPRA) or STARR-seq [49,52]. One caveat to such approach might be the inability of some enhancers to activate specific promoters [53]. In addition, these methods test putative enhancer sequences outside of their original genomic context. The modification of candidate enhancer sequences in their endogenous context can be accomplished through DNA-editing techniques such as CRISPR-Cas9 or TALEN-based methods [54–58]. However, compared to reporter constructs, these methods are not as precise in dissecting regulatory elements, and result in a large spectrum of indels at the target site and can in addition affect other genomic regions than the target site (off-target sites).

Another striking aspect of the candidate enhancers described in chapter 2 is their substantial overlap with Transposable Elements (TEs). Approximately 30% of the candidate enhancers overlapped with a TE, and most of these are even entirely within a TE. One well-known maize enhancer, that of *tb1* (candidate H1403), consists of a Hopscotch TE [59]. TEs are known to be a source of novel *cis*-regulatory elements as they require the transcriptional machinery of their host for gene expression [60–63]. To understand if enhancer sequences could originate from specific TEs, we looked for TE families showing enrichment for candidate enhancers. Analysis indicated three LTR Gypsy families (RLG00010, RLG00357, RLG01570) as significantly enriched for candidate enhancers. Comparison of enhancer-associated TEs with all TEs indicated a slightly longer size of the former, but no differences in terms of their distance to the nearest gene. In addition, candidates within TEs were distributed randomly within TE sequences, indicating the absence of overlap with promoters contained within TEs. Motif analysis of TE-associated enhancers revealed their enrichment for a site II motif (GGCCCA), but this motif was also enriched in candidate enhancers not overlapping with TE sequences. Future research has to show the exact function of the TE-associated enhancer candidates.

Validation and characterization of candidate enhancers

The identification of candidate enhancers is a first step in deciphering the regulatory potential of the genome but calls for validation of the predicted sequences. In this thesis, we aimed at validating candidate enhancers predicted by us and by others. For this, we generated stable transgenic maize lines containing GUS reporter constructs or Inverted Repeat (IR) silencing constructs and put effort in establishing a transient reporter assay (Chapter 3 and 4). Note that the GUS reporter system used in this study suffers from an unexpected high activity of the minimal *35S* promoter, which resulted in substantial "background" GUS expression in *min35S:GUS* transgenic lines, making it difficult to assess whether inclusion of putative enhancers increases the GUS expression levels (see Chapter 3 and 4). Nevertheless, our results from the different GUS transgenic lines indicate that some of the candidate sequences tested can act as transcriptional enhancers. Others can potentially act as transcriptional silencers or insulators, while some may act as enhancer in the one, but silencer/insulator in the other tissue examined.

Two candidate enhancers (H11/V426 and H112/V441) reduced GUS expression when included in the *min35S:GUS* transgene, suggesting a role as transcriptional silencer or insulator rather than transcriptional enhancer. Note that our reporter assay does not distinguish between silencer or insulator activity. A role for H112 as a silencer was supported by the downregulation of two

relatively closeby genes in the tissue in which they physically interacted with H112. These were the only genes that interacted with H112 in the bait region. However, genome-wide, the expression of genes physically interacting with H112 specifically in V2-IST tissue had a tendency to be higher in this tissue than in husk. In addition, we showed that two LUMR sequences that did not match the full set of criteria used to predict active enhancer candidates (ex-candidate V29 and neutral 4) were capable of acting as transcriptional enhancers in seedling tissues; other tissues than the one used to predict candidate enhancers (Chapter 4). One of these, neutral 4, in addition appeared to act as a silencer/insulator in ear tissues. This fits our hypothesis that the LUMRs identified in chapter 2 represent the genome-wide pool of intergenic *cis*-regulatory sequences.

In chapter 3, we focused on the suspected enhancer *Vgt1* and showed that *Vgt1* and its upstream region, coined *uva1,* can function separately, and also in combination, as tissue-specific transcriptional enhancers. No enhancing capacities were detected in the inner stem tissue and husk leaves, consistent with the absence of *Vgt1* and *uva1* from the list of candidate regions identified in Chapter 2. In husk leaves, *Vgt1*, *uva1* and the combined region appeared to act as a silencer or insulator. Further experiments indicated that silencing of *Vgt1* using IR constructs (*Vgt1*-IR lines) is sufficient to induce early flowering. This result is in line with previous studies associating sequence variations, including a TE insertion, at *Vgt1* with delays in flowering time [64–66]. In addition, we showed that *Vgt1* negatively impacts plant growth rate, as demonstrated by the faster growth rate of *Vgt1*-IR lines compared to non-transgenic plants. The effect on flowering time is in line with *Vgt1* regulating the downstream *AP2*-like floral repressor gene *ZmRap2.7*. In favour of this hypothesis, we show that the TSS of *ZmRap2.7* physically interacts with a region encompassing *Vgt1*. Evidence regarding *Vgt1* being a transcriptional enhancer of *ZmRap2.7* is still unclear though. In the tissues examined, we did not observe significant downregulation of *ZmRap2.7* expression in *Vgt1*-silenced lines, suggesting *Vgt1* does not regulate expression of *ZmRap2.7*.

Intriguingly, using Circular Chromosome Conformation Capture (4C), we identified genome-wide contacts of the *ZmRap2.7* TSS overlapping with genes and loci encoding long non-coding RNAs (lncRNAs), but also with five candidate enhancers in the bait chromosome. The latter contacts were mostly specific for V2-IST tissue, the tissue in which *ZmRap2.7* is expressed. We therefore hypothesized that *ZmRap2.7* might be under the regulation of additional regulatory sequences besides *Vgt1*, limiting the impact of only silencing the *Vgt1* regulatory sequence by IR constructs.

Table 1: Summary of the effects of sequences tested in the different tissues

Sequence Tested	V2 seedling	Ear tissues
Vgt1	enhancer	Silencer/insulator
uva1	enhancer	Silencer/insulator
uva1-Vgt1	enhancer	Silencer/insulator
H11	Silencer/insulator	Silencer/insulator
H112	Silencer/insulator	Silencer/insulator
V29	enhancer	Background level
N4	enhancer	Silencer/insulator
N7	Silencer/insulator	Background level

The results obtained from GUS transgenic lines for sequences indicated in table 1 were complicated by the presence of background GUS expression in *min35S:GUS* transgenic lines. The transgenic cassette incorporated into the genome of the transgenic lines contained, on top of the individual reporter constructs, the BAR selectable marker gene driven by the *mannopine synthase* promoter. Importantly, in tobacco the *MAS* promoter and subfragments thereof were shown capable of activating the minimal 35S promoter in a tissue-specific and wounding inducible manner [67,68]. We therefore suspect the *MAS* promoter to be responsible for the activation of the minimal 35S promoter resulting in background GUS expression in *min35S:GUS* transgenic lines. However, when taking the background GUS activity into account, most sequences tested seem capable of acting as regulatory sequences (Table 1). In addition, about half of these sequences appeared multi-functional: enhancing expression in one tissue, decreasing it in another. These results are in line with previous studies demonstrating a multi-functional potential of regulatory sequences in both animals and plants [69–74]. This multi-functional role of regulatory sequences is coherent with their enrichment for TFBS clusters [74–76] and with the existence of multi-functional TFs that can be activators or repressors [77–80]. In addition, the functionality of a regulatory sequence might be dependent on its genomic location and thereby genomic sequence context, which are flanking or relatively closeby sequences capable of functionally interacting with the sequence of interest. For instance, the LCR control region of the β-globin locus, normally acting as a strong

transcriptional enhancer, was shown to be sequence context specific and act as a transcriptional silencer in a different genomic location [81]. In addition, the fab-7 insulator from *Drosophila* can also act as a transcriptional enhancer in a different genomic context [82]. All together, these results demonstrate the complexity of transcriptional regulation and the necessity to consider regulatory sequences as a whole rather than fixed entities with one unique function.

Candidate enhancer interactome

Distant enhancers physically interact with their target gene(s) through formation of chromatin loops often referred to as Promoter-Enhancer (P-E) interactions (*a.o.* [83–88]). In addition, enhancers are known to interact with other regulatory regions including enhancers (E-E interactions), while promoters were shown to also interact with other promoters (P-P interactions) [88,89]. Note that also coding sequences can be included in interactions. Such interactions can be studied using Chromosome Conformation Capture (3C) based techniques. In this thesis, a detailed protocol to perform 3C, specifically in maize and Arabidopsis is reported (Chapter 5). Also, to better understand the type of contacts maize sequences are involved in, we used 4C, taking as baits the candidate enhancer H112/V441, neutral region 4 and the TSS of *ZmRap2.7.*

In line with previous studies, our results identified for all three baits P-E, E-E and P-P interactions. These include, for instance, interactions between the TSS of *ZmRap2.7* and other candidate regulatory sequences (E-P interactions) or coding regions (P-P interactions), and interactions between candidate H112/V441 and coding regions (E-P) or other candidate regulatory sequences (E-E). In human and mice cells it was shown that genes that are in physical contact with one another are often transcribed in a coordinated manner [89–92]. It would be interesting to see if genes interacting with the promoter of *ZmRap2.7* show a similar transcriptional behaviour.

We show that all three bait regions interact preferentially with other regions of the genome characterised by active chromatin features (open chromatin regions, LUMRs, H3K9ac enriched regions, candidate enhancers). Significant interaction with sequence regions displaying inactive features (high CHG methylation) was only detected with husk specific contacts (H contacts) or contacts common to both V2-IST and husk tissues (HV contacts) of bait H112/V441, but this might be due to the low number of contacts in these categories. The interactions of the two active bait regions with other active regions of the genome appear consistent with the activity status of the baits itself. In addition, the *ZmRap2.7* TSS bait was specifically associated with other candidate enhancers in V2-IST, a tissue in which *ZmRap2.7* is expressed, while bait

H112/V441 was interacting with other candidate enhancers in V2-IST but not in husk tissue. The genes with which H112/V441 interacts specifically in V2-IST tissue showed overall higher expression. These results are coherent with a recent Hi-C study in maize indicating that active regions interact with one another whether located within the same active chromatin compartment or different ones [93].

More work is needed to fully understand the characteristics of E-P interactions in maize, such as the number of target gene(s) per enhancer, the number of enhancer(s) per gene and the distance at which such interactions can take place. One intra-chromosomal E-P interaction reported in this study spanned 52 Mb, which is longer than the previously reported maximal distance for functional distal E-P interactions in animals (for *e.g.* [94]). The functionality of interactions identified in this thesis remains to be demonstrated, potentially through disrupting the sequence of interest using genome editing techniques [56,57], followed by monitoring effects on transcript levels at potential target genes or genome-wide [58]. Also note that in animals, the genome is organized into Topologically Associated Domains (TADs) that are stable across tissues and conserved among species [95,96]. TADs have an average size of 1 Mb, and for a large part restrict E-P interactions to within their limits [95,96]. These TADs are themselves contained into A and B compartments (active and inactive, respectively), and both intra- and inter-compartment interactions have been reported for A, but also B compartments [97,98]. In plants, the presence of TADs remains to be identified, suggesting A/B compartments are the main organizational structure of the genome [93]. In maize, this genomic organization into A/B compartments is complicated by the very repetitive nature of its genome with genes and regulatory elements being separated from one another by the presence of TE-rich regions. The maize genome organization increases the distances between *cis*-regulatory sequences and target genes therefore explaining the very long E-P distances reported in our study.

Conclusion

This thesis is a modest contribution to the identification and characterisation of plant regulatory sequences. On top of establishing a first genome-wide map of candidate regulatory sequences in the maize B73 reference genome, the work highlights similarities of plant and animal enhancers, but also indicates some of their differences. The thesis also gets into the details of some of the candidate regulatory sequences through a first characterisation of the *Vgt1* regulatory sequence, a few predicted candidate enhancers and control sequences. Finally, also a sneak peek into the 3D organization of the maize genome is given through identification of the chromatin interactions of regulatory sequences and a gene promoter.

References

1 Weber, B. *et al.* (2016) Plant Enhancers: A Call for Discovery. *Trends Plant Sci.* 21, 974–987
2 Marand, A.P. *et al.* (2017) Towards genome-wide prediction and characterization of enhancers in plants. *Biochim. Biophys. Acta* 1860, 131–139
3 Heintzman, N.D. *et al.* (2009) Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. *Nature* 459, 108–112
4 Creyghton, M.P. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* 107, 21931–21936
5 Tian, Y. *et al.* (2011) Global Mapping of H3K4me1 and H3K4me3 Reveals the Chromatin State-Based Cell Type-Specific Gene Regulation in Human Treg Cells. *PLoS ONE*
6 Oka, R. *et al.* (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* 18, 137
7 Rada-Iglesias, A. *et al.* (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283
8 Karmodiya, K. *et al.* (2012) H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* 13, 424
9 Du, Z. *et al.* (2013) Genome-Wide Analysis of Histone Modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in Oryza sativa L. Japonica. *Mol. Plant* 6, 1463–1472
10 Zhu, B. *et al.* (2015) Genome-Wide Prediction and Validation of Intergenic Enhancers in Arabidopsis Using Open Chromatin Signatures. *Plant Cell* DOI: 10.1105/tpc.15.00537
11 Zentner, G.E. *et al.* (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 21, 1273–1283
12 Bonn, S. *et al.* (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* 44, 148–156
13 Kim, T.-K. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187
14 Andersson, R. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461
15 Li, W. *et al.* (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* 17, 207
16 Hetzel, J. *et al.* (2016) Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12316–12321
17 Thurman, R.E. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature* 489, 75–82
18 Calo, E. and Wysocka, J. (2013) Modification of Enhancer Chromatin: What, How, and Why? *Mol. Cell* 49, 825–837
19 Daugherty, A.C. *et al.* (2017) Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. *Genome Res.* 27, 2096–2107
20 Maher, K.A. *et al.* (2017) Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* DOI: 10.1105/tpc.17.00581
21 Köhler, C. and Villar, C.B.R. (2008) Programming of gene expression by Polycomb group proteins. *Trends Cell Biol.* 18, 236–243
22 King, A.D. *et al.* (2016) Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep.* 17, 289–302
23 Pradeepa, M.M. (2016) Causal role of histone acetylations in enhancer function. *Transcription* 8, 40–47

24 Prabhakar, S. *et al.* (2006) Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* 16, 855–863
25 Pennacchio, L.A. *et al.* (2007) Predicting tissue-specific enhancers in the human genome. *Genome Res.* 17, 201–211
26 Visel, A. *et al.* (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40, 158–160
27 Cheng, Y. *et al.* (2014) Principles Of Regulatory Information Conservation Between Mouse And Human. *Nature* 515, 371–375
28 Pennacchio, L.A. and Visel, A. (2010) Limits of sequence and functional conservation. *Nat. Genet.* 42, 557–558
29 White, M.A. *et al.* (2013) Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11952–11957
30 Slattery, M. *et al.* (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39, 381–399
31 Criminisi, A. (2016) Machine learning for medical images analysis. *Med. Image Anal.* 33, 91–93
32 Shen, D. *et al.* (2017) Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248
33 Firpi, H.A. *et al.* (2010) Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinforma. Oxf. Engl.* 26, 1579–1586
34 Rajagopal, N. *et al.* (2013) RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* 9, e1002968
35 Capra, J.A. (2015) Extrapolating histone marks across developmental stages, tissues, and species: an enhancer prediction case study. *BMC Genomics* 16, 104
36 Kleftogiannis, D. *et al.* (2015) DEEP: a general computational framework for predicting enhancers. *Nucleic Acids Res.* 43, e6
37 Herman-Izycka, J. *et al.* (2017) Taking promoters out of enhancers in sequence based predictions of tissue-specific mammalian enhancers. *BMC Med. Genomics* 10, 34
38 Bu, H. *et al.* (2017) A new method for enhancer prediction based on deep belief network. *BMC Bioinformatics* 18, 418
39 Zhang, S. *et al.* (2017) pDHS-SVM: A prediction method for plant DNase I hypersensitive sites based on support vector machine. *J. Theor. Biol.* 426, 126–133
40 Zhang, S. *et al.* (2018) Prediction of DNase I hypersensitive sites in plant genome using multiple modes of pseudo components. *Anal. Biochem.* 549, 149–156
41 Banf, M. and Rhee, S.Y. (2017) Computational inference of gene regulatory networks: Approaches, limitations and opportunities. *Biochim. Biophys. Acta* 1860, 41–52
42 Qiu, Z. *et al.* (2016) Identification of Regulatory DNA Elements Using Genome-wide Mapping of DNase I Hypersensitive Sites during Tomato Fruit Development. *Mol. Plant* 9, 1168–1182
43 Schmitz, R.J. *et al.* (2013) Patterns of population epigenomic diversity. *Nature* 495, 193–198
44 Kawakatsu, T. *et al.* (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat. Plants* 2, 16058
45 Andersson, R. (2015) Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* 37, 314–323
46 Raab, J.R. and Kamakaka, R.T. (2010) Insulators and promoters: closer than we think. *Nat. Rev. Genet.* 11, 439–446
47 Palstra, R.-J. and Grosveld, F. (2012) Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Front. Genet.* 3,

48 Kolovos, P. *et al.* (2012) Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* 5, 1
49 Kheradpour, P. *et al.* (2013) Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* 23, 800–811
50 Qi, H. *et al.* (2015) Functional validation of a constitutive autonomous silencer element. *PloS One* 10, e0124588
51 Young, R.S. *et al.* (2017) Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome Biol.* 18, 242
52 Arnold, C.D. *et al.* (2013) Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* 339, 1074–1077
53 Zabidi, M.A. *et al.* (2015) Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559
54 Christian, M. *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761
55 Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
56 Sprink, T. *et al.* (2015) Plant genome editing by novel tools: TALEN and other sequence specific nucleases. *Curr. Opin. Biotechnol.* 32, 47–53
57 Bortesi, L. and Fischer, R. (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52
58 Osterwalder, M. *et al.* (2018) Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239
59 Studer, A. *et al.* (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163
60 Santangelo, A.M. *et al.* (2007) Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. *PLoS Genet* 3, e166
61 Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405
62 Ito, H. *et al.* (2013) Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. *Gene* 518, 256–261
63 Chuong, E.B. *et al.* (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86
64 Salvi, S. *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci.* 104, 11376–11381
65 Bolduc, N. *et al.* (2012) Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes Dev.* 26, 1685–1690
66 Castelletti, S. *et al.* (2014) A MITE Transposon Insertion Is Associated with Differential Methylation at the Maize Flowering Time QTL Vgt1. *G3 Genes Genomes Genet.* 4, 805–812
67 Guevara-García, A. *et al.* (1993) Tissue-specific and wound-inducible pattern of expression of the mannopine synthase promoter is determined by the interaction between positive and negative cis-regulatory elements. *Plant J. Cell Mol. Biol.* 4, 495–505
68 Guevara-García, A. *et al.* (1999) The mannopine synthase promoter contains vectorial cis-regulatory elements that act as enhancers and silencers. *Mol. Gen. Genet. MGG* 262, 608–617
69 Leah, R. *et al.* (1994) Identification of an enhancer/silencer sequence directing the aleurone-specific expression of a barley chitinase gene. *Plant J. Cell Mol. Biol.* 6, 579–589
70 Schaeffer, H.J. *et al.* (1995) Identification of enhancer and silencer regions involved in

salt-responsive expression of Crassulacean acid metabolism (CAM) genes in the facultative halophyte Mesembryanthemum crystallinum. *Plant Mol. Biol.* 28, 205–218
71 Aza-Blanc, P. *et al.* (1997) Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 89, 1043–1053
72 Fitzpatrick, G.V. *et al.* (2007) Allele-Specific Binding of CTCF to the Multipartite Imprinting Control Region KvDMR1. *Mol. Cell. Biol.* 27, 2636–2647
73 Evans, N.C. *et al.* (2012) Sparkling insights into enhancer structure, function, and evolution. *Curr. Top. Dev. Biol.* 98, 97–120
74 Levo, M. and Segal, E. (2014) In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* 15, 453–468
75 Gotea, V. *et al.* (2010) Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* 20, 565–577
76 Chen, H. *et al.* (2015) An integrative analysis of TFBS-clustered regions reveals new transcriptional regulation models on the accessible chromatin landscape. *Sci. Rep.* 5, 8465
77 Brembeck, F.H. *et al.* (2000) Dual function of the epithelial specific ets transcription factor, ELF3, in modulating differentiation. *Oncogene* 19, 1941–1949
78 Ou, X.-M. *et al.* (2004) Dual functions of transcription factors, transforming growth factor-beta-inducible early gene (TIEG)2 and Sp3, are mediated by CACCC element and Sp1 sites of human monoamine oxidase (MAO) B gene. *J. Biol. Chem.* 279, 21021–21028
79 Parker, D.S. *et al.* (2011) The cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity. *Sci. Signal.* 4, ra38
80 Li, T. *et al.* (2016) A Dual-Function Transcription Factor, AtYY1, Is a Novel Negative Regulator of the Arabidopsis ABA Response Network. *Mol. Plant* 9, 650–661
81 Feng, Y.-Q. *et al.* (2005) The human beta-globin locus control region can silence as well as activate gene expression. *Mol. Cell. Biol.* 25, 3864–3874
82 Rodin, S. *et al.* (2007) New properties of Drosophila fab-7 insulator. *Genetics* 177, 113–121
83 Tolhuis, B. *et al.* (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465
84 Louwers, M. *et al.* (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. *Plant Cell Online* 21, 832–842
85 Sanyal, A. *et al.* (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109
86 Chepelev, I. *et al.* (2012) Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* 22, 490–503
87 Cao, S. *et al.* (2014) A Distal CCAAT/NUCLEAR FACTOR Y Complex Promotes Chromatin Looping at the FLOWERING LOCUS T Promoter and Regulates the Timing of Flowering in Arabidopsis. *Plant Cell Online* 26, 1009-1017. DOI: 10.1105/tpc.113.120352
88 Mumbach, M.R. *et al.* (2016) HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922
89 Li, G. *et al.* (2012) Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* 148, 84–98
90 Schoenfelder, S. *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61
91 Papantonis, A. *et al.* (2012) TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.* 31, 4404–4414
92 Fanucchi, S. *et al.* (2013) Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606–620
93 Dong, P. *et al.* (2017) 3D Chromatin Architecture of Large Plant Genomes

Determined by Local A/B Compartments. *Mol. Plant* 10, 1497–1509
94 Lettice, L.A. *et al.* (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735
95 Dixon, J.R. *et al.* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336
96 de Laat, W. and Duboule, D. (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506
97 Lieberman-Aiden, E. *et al.* (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293
98 Rowley, M.J. *et al.* (2017) Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* 67, 837-852.e7