The packaging status of chromatin is directly associated with local or global accessibility of underlying DNA to trans-acting factors which ultimately affects gene expression. Open chromatin is referred to chromatin with open conformation, facilitating changes at the underlying accessible cis-regulatory elements [Bell et al., 2011], and is generally associated with regulatory regions that are nucleosome-free or depleted. Open chromatin exhibits high sensitivity to any DNA shearing processes, including (a) physical shearing, such as sonication, (b) biochemical shearing through enzymes, including endonuclease deoxyribonuclease I (DNase I) which cuts single- and double-stranded DNA and chromatin and releases products with 5′-phosphorylated and 3′-hydroxylated ends without any sequence preference, endonuclease deoxyribonuclease II (DNase II) which hydrolyzes DNA under acidic conditions and yields products with 3′-phosphates, micrococcal nuclease (MNase) which preferentially cleaves linker DNA between 2 nucleosomes, and other restriction enzymes [Todd and Harrison, 1979; Wu, 1980; Keene et al., 1981; McGhee et al., 1981; Kaye et al., 1984; Scott et al., 1984], and (c) chemical cleavage using reagents such as methidiumpropyl-EDTA, iron (II) (MPEiFe (II))
which cleaves double-helical DNA with relative lower sequence specificity, and orthophenanthroline-cuprous complex which preferentially recognizes and cleaves the linker between nucleosomes resulting in the release of nucleosomes [Cartwright and Elgin, 1982; Cartwright et al., 1983].

The hypersensitivity of open chromatin to DNase I has been extensively researched. A genomic region exhibiting increased sensitivity to DNase I cleavage is referred to as a DNase I hypersensitive site (DHS). DHSs have been studied since the 1970s [Scott and Wigmore, 1978; Varshavsky et al., 1978; Weisbrod and Weintraub, 1979; Gross and Garrard, 1988] and were first discovered in chromatin isolated from simian virus 40 (SV40) [Scott and Wigmore, 1978; Varshavsky et al., 1978; Gross and Garrard, 1988]. The existence of DHSs in eukaryotic chromatin was first detected in *Drosophila* [Wu, 1980; Keene et al., 1981], globin genes in chicken and humans [Stalder et al., 1980; Groudine et al., 1983], and the metallothionein gene and several other genes in mouse [Mather and Perry, 1983; Senear and Palmiter, 1983; Krebs and Peterson, 2000]. This technique was used to map DHSs associated with heat shock genes in *Drosophila* [Wu, 1980; Keene et al., 1981], globin genes in chicken and humans [Stalder et al., 1980; Groudine et al., 1983], and the metallothionein gene and several other genes in mouse [Mather and Perry, 1983; Senear and Palmiter, 1983; Bender et al., 2000; Bulger et al., 2003]. Similarly, DHSs across approximately 300 kb of the human genome were profiled using quantitative PCR assays (qPCR). These DHSs were also found to be associated with classical *cis*-regulatory functions, such as promoters, enhancers, locus control regions, insulators, and other novel elements [Pfeifer and Riggs, 1991; Feng and Villedo, 1999; McArthur et al., 2001; Dorschner et al., 2004; Martins et al., 2007]. DHSs distributed within the TAL1/SCL locus (T-cell acute lymphocytic leukemia-1), also known as SCL (stem cell leukemia) between human embryonic stem cells and K562 cells, were successfully mapped using real-time PCR analysis [Follows et al., 2007].
High-Throughput Methodologies for Identification of DHSs

The Southern blot-based DHS assay is time-consuming and low throughput. This method is only suitable for analysis of a single locus or short stretch of DNA sequences. With the advent of high throughput sequencing techniques, several methods have been developed to identify and map DHSs across an entire genome. Among these methods, DNase I digestion followed by sequencing (DNase-seq) [Boyle et al., 2008] has been applied for identification of DHSs in several species. This method includes isolation of nuclei, DNase I digestion, end-blunting of DNase I digested high molecular weight DNA, sequential ligation with 2 different adaptors, and Illumina sequencing after PCR enrichment of ligated DNA fragments. DHS maps have been developed in several model animal species with sequenced genomes, such as human, mouse, and Drosophila [Crawford et al., 2004, 2006a, b; Sabo et al., 2004, 2006; Weil et al., 2004; Boyle et al., 2008; Song and Crawford, 2010]. In addition to DHS mapping, formaldehyde-assisted isolation of regulatory elements (FAIRE) [Giresi et al., 2007], chromatin immunoprecipitation (ChIP) [Johnson et al., 2007], MNase-seq [Schones et al., 2008], Sono-seq [Auerbach et al., 2009], NOMe-seq [You et al., 2011], MeDIP footprint assay [Bell et al., 2010], and ATAC-seq [ Buenrostro et al., 2013] have also provided supplemental information to indicate the presence of open chromatin. For example, FAIRE is a method to isolate and identify nucleosome-depleted regions in the genome which was initially established in yeast and has also been applied to human [Giresi et al., 2007; Hogan et al., 2006] and maize [Louwers et al., 2009]. Similar to DNase-seq, FAIRE appears to efficiently map all CREs. DNase-seq and FAIRE have been used as complementary methods for identifying open chromatin in human [Song et al., 2011].

DNase-seq, FAIRE-seq, and ChIP-seq are the most commonly used methodologies for identification of open chromatin. Each method has its own advantages and disadvantages. DNase-seq is the gold standard in identifying all CREs, but it can’t distinguish between active and poised DNA elements. FAIRE-seq cannot distinguish permanent interaction from transient interaction between DNA and trans-acting factors after formaldehyde fixation. ChIP-seq requires a specific antibody recognizing a specific trans-acting factor bound to various genomic loci. Similarly, Sono-seq needs chromatin cross-linking and sonication for chromatin fragmentation which has limitations similar to FAIRE-seq. NOMe-seq offers the ability to profile both nucleosome occupancy and CpG methylation on the same DNA strand, but it depends on a methyltransferase which in vitro marks open chromatin containing GC methylation. Thus, any factor affecting the accessibility of the enzyme to the DNA will mask open chromatin identification.

Several other methods have been developed for identifying higher-order chromatin structure and dynamics, including 3C (chromosome conformation capture) [Dekker et al., 2002], 4C (circular chromosome conformation capture and chromosome-conformation-capture-on-chip) [Zhao et al., 2006], 5C (chromosome conformation capture carbon copy) [Dostie and Dekker, 2007], 6C (combined chromosome conformation capture ChIP cloning) [Tiwari et al., 2008], Hi-C [Lieberman-Aiden et al., 2009], and ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) [Fullwood et al., 2009]. These methods can be applied to study long-range chromatin interactions that occurred locally or globally within the genome and can be potentially used for studying chromatin accessibility [Sajan and Hawkins, 2012].

Genomic Locations of DHSs

A relative high DHS-positive rate associated with some silenced genes was observed in HeLa S3 cells which either suggests that chromatin related to inactive genes still retains a certain degree of openness and accessibility [Wang et al., 2012] or indicates that those DHS-associated silenced genes are poised for expression through RNA polymerase II binding in a development dependent manner [Muse et al., 2007; Zeitlinger et al., 2007]. DHSs, however, are generally not associated with inactive genes. The presence of DHSs is thus viewed as a method for identification of active genes or genes poised for tissue-specific expression [Krebs and Peterson, 2000]. The main types of DHSs within cell types or tissues fall into the following 4 categories: constitutive, inducible, tissue specific, and development specific [Gross and Garrard, 1988]. Constitutive and inducible DHSs are usually found in gene promoter regions. The vast majority of well described DHSs are associated with functional genomic elements [Eisenberg et al., 1985; Gross and Garrard, 1988; Krebs and Peterson, 2000; Cockerill, 2011], including enhancers [McGinnis et al., 1983; Nelson and Groudine, 1986], silencers [Fritton et al., 1983; Shore et al., 1987], promoters [Wu et al., 1979; Staldet al., 1980; Wu, 1980; Ferl and Nick, 1987], terminators [Almer et al., 1986; Gargiulo et al., 1985; Yagi et al., 1986], upstream activation sequences [Lohr, 1984], replication origins [Varshavsky et al., 1979; Borchsenius et al., 1981; Fagrelius and Livingston, 1980], locus control regions [Grosveld et al., 1987],...
DHSs in the Human Genome

DHS mapping has been used as a powerful tool to identify CREs in the human genome. It is one of the most important methodologies used in the ENCODE (Encyclopedia of DNA Elements) project in human [Birney et al., 2007; Bernstein et al., 2012] and other model animals [Gerstein et al., 2010; Roy et al., 2010; Stamatoyannopoulos et al., 2012] which has been proved to lead toward a comprehensive and functional annotation of genomic sequences, especially for previously incompletely annotated ones [Neph et al., 2012; Thurman et al., 2012; Vernot et al., 2012].

Most human DHSs are localized within intergenic regions, 5′-UTRs, 3′-UTRs, introns, and exons of active genes. These genomic regions are known to be associated with various CREs [Boyle et al., 2008]. Thus, CREs can be identified by the presence of DHSs together with other chromatin signatures. For example, enhancers can be identified by DHSs together with P300/CBP binding and histone modifications marks H3K4me1 and H3K27ac [Heintzman et al., 2007, 2009; Ong and Corces, 2011; Zentner et al., 2011; Kolovos et al., 2012]. CTCF (CCCTC-binding factor) is the only insulator protein identified in vertebrates thus far [Valenzuela and Kamakaka, 2006; Wallace and Felsenfeld, 2007]. Correlation analysis between DNase-seq and CTCF-based ChIP-seq datasets provided strong evidence that DHSs overlapping with CTCF binding sites can potentially function as insulators across the human genome, although this is not universally true as not all insulators/silencers are bound by CTCF [Wang et al., 2012].

Functionally, DHSs have been found to be closely associated with various types of biological functions, including gene expression and regulation, nucleosome stability, replication, and cohesion [Parelho et al., 2008; Shu et al., 2011; Song et al., 2011; Wang et al., 2012; Winter et al., 2013]. Most importantly, DHSs were found to be associated with phenotypic variations in humans, including human diseases [Schaub et al., 2012], such as lung disease [Bischof et al., 2012] and Type 2 diabetes [Stitzel et al., 2010].

Identification of DHSs in Plants

The local presence of DHSs within a few plant genes was identified using DNase I cleavage coupled with Southern blot hybridization, including the Adh gene in maize and Arabidopsis [Wu et al., 1979; Vega-Palas and Ferl, 1995], ribosomal RNA genes (rRNA) in barley [Dimitrova et al., 2009] and wheat [Thompson and Flavell, 1988], rbcS in pea [Gorz et al., 1988], heat-shock and abscisic acid inducible genes in wheat [Loer and Spiker, 1992], Protease inhibitor I in tomato [Conconi and Ryan, 1993], and HSP 70A and RBCS2 in Chlamydomonas [Lodha and Schroda, 2005]. Remarkably, this Southern blot-based method was used to identify all DHSs within an 80-kb genomic region in Arabidopsis [Kodama et al., 2007a, b]. Local DHSs were also detected by DNase I digestion followed by locus-specific PCR [Shu et al., 2013]. High-throughput techniques have recently been developed in plants for genome-wide mapping of DHSs. These methods coupled partial DNase I cleavage with either next generation sequencing (DNase-seq) [Zhang et al., 2012a, b] or DNA microarray analysis (DNase-chip) [Shu et al., 2012, 2013].

The fact that cell walls and chloroplasts are present in plant cells requires an adaptation of the DNase-seq method developed for cultured cells in humans for successful application in plants. First, caution is required during nuclei isolation and purification to avoid degradation of plant nuclei/chromatin. Since sequencing reads generated from the DNase I cut sites are used to identify DHSs in DNase-seq, broken sites resulting from degradation will lead to false positive mapping results. To minimize unspecific damage to nuclei or chromatin, sucrose needs to be added to the nuclei isolation and washing buffer to stabilize the nuclei; centrifugation should be carried out at the lowest speed possible, a reduced concentration of Triton X-100 should be used in washing buffer, and washes should be kept to a minimum. Secondly, we need to thoroughly disrupt cell walls and other fibrous structures by grounding frozen plant tissues into a fine powder using liquid nitrogen. Fine grinding will also ensure the maximum yield of nuclei from a certain amount of tissue which sometimes can be a limiting factor for experiments. Lastly, we need to remove as much chloroplast contamination as possible by adding 0.4–0.5% Triton X-100 to the washing buffer and performing 2 or 3 washes until a white/yellow pellet is clearly visible. Most importantly, certain concentrations of Triton X-100 can destabilize the membrane of chloroplasts [Saxena et al., 1985]. A high concentration of Triton X-100, however, not only degrades the membrane of chloroplasts to eliminate chloro-
plast contamination but also damages outer nuclear membranes and makes nuclei more fragile which consequently results in unspecific degradation of nuclei/chromatin [Sikorskaite et al., 2013]. Thus, concentration of Triton X-100 should be optimized depending on which tissue and species are being used for nuclei isolation.

**DHSs across Model Plant Genomes**

*Comparison of DHS Distribution in Plants and Humans*

DHSs have been identified across the entire genomes in both rice and *Arabidopsis* using DNase-seq [Zhang et al., 2012a, b]. Association analyses between DNase-seq and RNA-seq datasets revealed that the sensitivity of DHSs is positively correlated with the expression level of genes associated with the DHSs. DHSs were found to be more frequently associated with highly expressed genes than weakly expressed or silent genes and were associated with various types of functional DNA elements, such as the conserved noncoding sequences in grass species [Zhang et al., 2012b]. DHSs were also found to be hotspots for recruiting cytoplasmic organelle DNA in both plant and human [Wang and Timmis, 2013]; for example, chloroplast DNA preferentially inserts into rice DHSs, while mitochondrial DNA tends to insert more frequently into DHSs in human.

Genome-wide DHS datasets developed in rice and *Arabidopsis* allowed us to compare the distribution of DHSs in plants vs. humans. Among all DHSs analyzed, about 39% of DHSs are associated with introns in the human genome which is remarkably more than those in both rice (11%) and *Arabidopsis* (5%), respectively (fig. 2), indicating that considerably more regulatory elements target the introns in the human genome than in rice and *Arabidopsis*. This difference is most likely caused by the distinct difference in gene structure between human and plants. For example, the average intron size in human is about 6,155 bp which is significantly longer than the average 398- and 180-bp-long intron size in rice and *Arabidopsis* genes, respectively. In addition, total intron sequences cover about 76% of human genes which is also remarkably more than the roughly 38 and 30% of coverage found in rice and *Arabidopsis* genes, respectively. Due to a more compact genome, with a greater gene density and a lower percentage of intergenic sequences in *Arabidopsis* as compared to rice and humans, *Arabidopsis* has more DHSs present in putative promoter regions and less DHSs located in intergenic regions (fig. 2).

**Tissue-Specific DHSs**

It is well known that tissue-specific gene expression plays an essential role in tissue/cell identity which is most likely directed by tissue/cell-specific regulatory elements. Identification of tissue-specific CREs is not only a fundamental task in developmental biology but also paves the way to uncover the molecular mechanism of tissue differentiation and development. DNase-seq is an efficient and powerful approach for identifying dynamic changes in tissue/cell-specific DHSs in both human [Song et al., 2011] and plant genomes [Zhang et al., 2012a, b]. Among ~870,000 total DHSs across 7 different types of human cell lines, only 30–40% of DHSs overlap between any 2 cell types analyzed, indicating the majority of DHSs function in cell/tissue-type identity by regulating gene activities [Song et al., 2011]. Similarly, tissue-specific DHSs were also detected in both rice and *Arabidopsis*. Between 155,025 and 97,975 DHSs identified from rice callus and seedling tissues, respectively, there is only 49% (76,057) overlap between callus and seedling DHSs. Thus, approximately 51% of callus DHSs function in a tissue specific manner and are likely essential for initiation and maintenance of rice callus tissue [Zhang et al., 2012a]. In *Arabidopsis*, 24% (9,926 of 41,000) flower DHSs are flower specific, and 22% (8,520 of 38,000) leaf DHSs are unique in leaf tissue [Zhang et al., 2012b]. Further analysis showed that genes regulated by flower-specific DHSs mainly function in developmental regulation of flowers, embry-
os, seeds, and fruits, supporting that tissue-specific DHSs function in the specification of corresponding tissues. For genomic locations, a high proportion of DHSs was detected in transcription start sites (TSSs) and intergenic regions of cell-type specific genes in human [Natarajan et al., 2012]. We also found DHSs enriched in TSS regions of either up-regulated or down-regulated genes in rice seedlings or calli [Zhang et al., 2012a]. Similarly, DHSs also significantly enriched around TSSs and 5′-UTRs of genes specific for Drosophila embryo development [Thomas et al., 2011].

**DNA Footprints Associated with DHSs in Plants**

Functions of CREs are determined by their interactions with regulatory proteins. The DNA sequences bound by regulatory proteins are shielded from nuclease cleavage as compared to immediately flanking regions, therefore resulting in the presence of DNA footprints [Galas and Schmitz, 1978]. Footprints of protein binding have been globally identified in both yeast and human using datasets derived from DNase-seq [Hesselberth et al., 2009; Boyle et al., 2011]. Similar analyses conducted in Arabidopsis demonstrated the power of DHS datasets in revealing protein-binding footprints [Zhang et al., 2012a]. The MADS-domain transcription factors APETALA1 (AP1) and SEPALLATA3 (SEP3) play key roles in regulating flower development in A. thaliana and are among the best characterized transcription factors in plants. Genome-wide AP1- and SEP3-binding sites were generated by ChIP-seq experiments [Kaufmann et al., 2009, 2010]. A total of 1,942 AP1- and 4,281 SEP3-binding sites were identified in flower tissue. We found that 1,843 (94.9%) of the AP1-binding sites and 3,841 (89.7%) of the SEP3-binding sites overlapped with DHSs identified in flower tissue. In addition, the ChIP-seq signal peaks of most AP1- and SEP3-binding sites overlapped with the DNase-seq signal peaks of the corresponding DHSs [Zhang et al., 2012a]. These results show that AP1- and SEP3-binding sites are well covered by DHSs, demonstrating the universal DNase I sensitivity associated with all CREs. In addition, the position of the CC[A/T]_GG motif specific for SEP3 binding was clearly less digested by DNase I compared to the flanking sequences which is most likely caused by the protection afforded by SEP3 binding [Zhang et al., 2012a]. Lastly, the same approach was extended to identify transcription factor binding footprints of a given genomic region [Zhang et al., 2012a, b]. For example, due to protection by MADS transcription factor binding, a clear footprint corresponding to a MADS box motif is only present in the promoter of the SUP gene from flower tissue but completely absent in the same locus from leaf tissue. These results showed that deep and high-quality DHS datasets can be used to confirm if a footprint is associated with the predicted or putative protein binding.

**Epigenetic Features of DHSs**

DHSs represent nucleosome depleted/dynamic regions in both human and plants, but well phased nucleosomes were clearly observed around all DHSs of the rice genome [Zhang et al., 2012b]. A distinct enrichment of H3K27me3, however, was observed within DHSs in the intergenic regions of the rice genome. Similarly, H3K27me3 and H3K9me2/3 modifications are enriched in distal DHSs in human [Boyle et al., 2008] which mostly are located within intergenic regions. The biological role of DHSs enriched with H3K27me3 need to be further elucidated, although they possibly function in tissue specificity through regulating expression of tissue-specific genes in both plant and human.

At the underlying DNA level, local DNase I-sensitive regions with reduced DNA methylation have been reported in pea, barley, and corn chromatin [Klaas and Amasino, 1989]. At the genome wide level, all DHSs in rice show a lower level of DNA methylation than the flanking DNA sequences. Similar hypomethylation trends were also observed in seedling and callus specific DHSs within the genes and the 200-bp downstream regions of the genes. However, the methylation level of tissue-specific DHSs in promoters for rice is significantly elevated relative to the overall methylation level of all seedling DHSs within promoters, suggesting that differential DNA methylation in promoters of tissue-specific DHSs may play a key role in tissue-specific gene expression [Zhang et al., 2012a]. In the Populus trichocarpa genome, a negative correlation was observed between promoter and/or gene body methylation and tissue-specific gene expression levels [Vining et al., 2012]. Among 30,679 annotated genes in shoot apical meristematic (SAM) tissue, methylated DHSs were found to be strongly enriched within 26% promoter regions and 55% gene bodies (exons) [Lafon-Placette et al., 2013] which is significantly higher than the overall methylation levels in the poplar genome (17% for promoters and 15% for exons) [Vining et al., 2012]. These results indicate that differentially methylated DHSs within both promoters and exons are associated with tissue-specific gene expression in poplar.
Plant genome research has entered the post-genome sequencing era. Sequencing of the entire genomes of plant species has become routine. Genes in a sequenced genome can now be readily identified and annotated by the powerful RNA-seq technique. However, the expression and regulation of genes in eukaryotes are controlled by an orchestrated binding of regulatory proteins to various types of CREs. Identification of CREs, especially those located outside of promoter regions, such as enhancers, has proven to be highly challenging. Thus, identification and characterization of CREs will be one of the future frontiers in plant genome research. DHS is a universal mark for all active CREs, thus, DHS provides a powerful tool for future CRE research.

Environmental abiotic stresses, such as drought, cold, heat, and high salinity, can induce various biochemical and physiological changes in plants which greatly affect plant growth and development. Understanding the mechanisms for stress adaptation and tolerance is one of the most important and challenging goals in plant sciences and holds the key for future crop improvement. Growing evidence has indicated that epigenetic mechanisms are involved in the regulation of expression of stress-inducible genes [Chinnusamy and Zhu, 2009; Kim et al., 2010; Mirouze and Paszkowski, 2011]. Thus, abiotic stresses can alter chromatin structure which will ultimately impact the binding of regulatory proteins to CREs and affect the expression of the genes associated with CREs. Thus, we expect that DHS mapping will be a powerful tool to reveal CREs associated with stress response and to identify genes that can be manipulated for crop stress tolerance.

Acknowledgments

DHS research in our lab has been supported by grant (DBI-0923640) from the National Science Foundation.


Varshavsky AJ, Sundin O, Bohn M: A stretch of ‘late’ SV40 viral DNA about 400 bp long which includes the origin of replication is specifically exposed in SV40 minichromosomes. Cell 16:453–466 (1979).


