The Genetic Map Enters Its Second Century

Steven Henikoff

Howard Hughes Medical Institute and Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

The Genetics Society of America (GSA) Medal is awarded to an individual for outstanding contributions to the field of genetics in the past 15 years. Recipients of the GSA Medal are recognized for elegant and highly meaningful contributions to modern genetics and exemplify the ingenuity of GSA members. The 2015 recipient is Steven Henikoff, whose achievements include major contributions to *Drosophila* genetics and epigenetics, *Arabidopsis* genetics and epigenetics, population and evolutionary genetics, genomic technologies, computational biology, and transcription and chromatin biology. Among these achievements, Henikoff elucidated the mechanism for position-effect variegation, revealed a central role for variant histones in nucleosome assembly at active genes, and provided new insights into genome evolution. He has also developed widely used computational tools for genome and protein analysis and new strategies for mapping chromatin-binding sites.

Genetics was born with Gregor Mendel's classic paper (Mendel 1866), but it was A. H. Sturtevant's introduction of the first genetic map in 1913 that defined the field of genetics as we know it today (Sturtevant 1913). For several decades, the genetic map was an ordering of genes along the chromosome determined by crossing over and chromosomal rearrangement. Beginning in the 1970s, restriction endonucleases allowed for genetic mapping to be based directly on DNA, which was eventually supplanted by DNA sequencing. As we enter the second century of genetic mapping, we are witnessing a progression from a genetic map defined by DNA sequence to a map enriched by the epigenomic landscape. Here I describe this latest transformation of the genetic map from the perspective of an observer and participant.

> As we enter the second century of genetic mapping, we are witnessing a progression from a genetic map defined by DNA sequence to a map enriched by the epigenomic landscape.
>
> — S.H.

My interest in genetic maps began in the mid-1980s when I discovered a glaring exception to Sturtevant's linear order of genes along the chromosome: a pair of "nested genes," in which a *Drosophila* pupal cuticle protein is encoded within an intron of the *Gart* purine biosynthetic pathway gene encoded on the opposite strand (Henikoff *et al.* 1986). By then, traditional "forward" genetics based on a genetic map was giving way to "reverse" genetics, which begins with homology searching and ends with phenotypic analysis of a mutant. Much of our attention during the reverse-genetics era was focused on improving homology searches, leading to popular tools such as the BLOSUM series of amino acid substitution matrices (Henikoff and Henikoff 1992) and SIFT (Sorting Intolerant From Tolerant) by graduate student Pauline Ng for predicting deleterious substitutions (Ng and Henikoff 2001). Homology searching led to discovery of a novel *Arabidopsis* DNA methyltransferase ("chromomethylase") (Henikoff and Comai 1998), along with a decade-long collaboration with Luca Comai and a thesis project for graduate student Claire McCallum. Confronted with the need to obtain chromomethylase mutants, Claire came up with the idea of screening PCR-amplified pools of mutagenized DNA for induced point mutations (McCallum *et al.* 2000). What we called "TILLING" (Targeting Induced Local Lesions IN Genomes) evolved into a reverse-genetic service for the *Arabidopsis* (Till *et al.* 2003), maize (Till *et al.* 2004), and *Drosophila* (Cooper *et al.* 2008) communities based on mismatch cleavage discovery of point mutations. Luca Comai and his colleagues have since developed TILLING into a sequencing-based strategy for efficient discovery and mapping of mutations and polymorphisms in crop plants (Tsai *et al.* 2011). TILLING for nontransgenic crop improvement also continues in the developing world, an effort spearheaded by former Seattle TILLING Project leader

Bradley Till at the Food and Agricultural Organization of the United Nations.

In the late 1990s postdoc Bas van Steensel had the idea of creating genome-wide epigenetic maps using tethered Dam DNA methyltransferase to mark sites of DNA- and chromatin-binding proteins along the chromosome (van Steensel and Henikoff 2000; van Steensel et al. 2001). Later in his own lab, Bas developed this DamID technique into a tool for probing the large-scale structure of the nucleus, mapping, for example, the five "colors" of Drosophila chromatin (Filion et al. 2010). Meanwhile, our introduction of microarray-based DNA methylation profiling as a readout platform for DamID led to the discovery of gene-body methylation by postdocs Robert Tran and Daniel Zilberman (Tran et al. 2005). Daniel showed in his own lab at the University of California at Berkeley that this epigenomic feature is ancestral for eukaryotes (Zemach et al. 2010). As readout technologies advanced, epigenomic mapping flourished, including Solexa-based mapping of DNA methylation imprints in the Arabidopsis endosperm by postdoc Mary Gehring (Gehring et al. 2009).

Dr. Henikoff's research has moved the entire field of genetics forward through a combination of technical innovations and fundamental discoveries. His selection as the recipient of the GSA Medal is a fitting honor to a scientist who inspires so many of us in so many different fields.

—Dan Gottschling, Fred Hutchinson Cancer Research Center

Our most recent tool development efforts have been motivated by a long-term interest in understanding epigenetic phenomena. Postdoc Kami Ahmad's discovery of the replication-independent pathway for nucleosome assembly that incorporates the H3.3 histone variant (Ahmad and Henikoff 2002) led graduate student Yoshiko Mito to introduce a biotin-tagging protocol for chromatin immunoprecipitation (ChIP) of H3.3-containing nucleosomes (Mito et al. 2005, 2007). Our genome-wide implementation of the classical salt-fractionation method indicated that the low-salt-soluble "active" chromatin fraction is enriched for the H2A.Z replication-independent histone variant and so might be turning over during transcription (Henikoff et al. 2009). To directly test this possibility, postdoc Roger Deal introduced the CATCH-IT (Covalent Attachment of Tags to Capture Histones and Identify Turnover) metabolic labeling method for measuring nucleosome turnover (Deal et al. 2010). Based on CATCH-IT we estimated that nucleosomes over epigenetic regulatory elements were turning over too fast for histone modifications themselves to be inherited between cell divisions. Graduate student Sheila Teves used CATCH-IT to show that transcription drives nucleosome turnover (Teves and Henikoff 2011), and postdoc Fan Yang used it to show that anthracycline chemotherapeutic drugs enhance transcription-coupled turnover (Yang et al. 2013) and double-strand breaks (Yang et al. 2015) around mouse promoters. Roger Deal's interest in applying high-resolution epigenomic tools to the study of Arabidopsis root development led him to introduce INTACT (Isolation of Nuclei Tagged in specific Cell Types) for cell-type-specific affinity purification of nuclei (Deal and Henikoff 2010). INTACT has also been applied to animal model systems, including worms, by postdoc Florian Steiner, and flies, by us and others (Henry et al. 2012; Steiner et al. 2012).

Our interests in both histone variants and DNA methylation converged with Daniel Zilberman's discovery that H2A.Z and DNA methylation are mutually exclusive in Arabidopsis (Zilberman et al. 2008), followed by evidence from graduate student Melissa Conerly that this is also the case for gene bodies in cancer cells (Conerly et al. 2010). Studies of Drosophila H2A.Z by graduate student Chris Weber showed that just downstream of the transcriptional start site nucleosomes are preferentially "homotypic," containing two H2A.Z molecules (Weber et al. 2010), which suggested that disruption of nucleosomes by RNA Polymerase II (RNAPII) facilitates H2A.Z incorporation. Chris next asked how the collision of RNAPII with a nucleosome affects RNAPII transit, but he realized that published methods for RNAPII profiling in higher eukaryotes were inadequate. He reasoned that the extreme insolubility of RNAPII would allow recovery of its attached nascent transcript in the washed chromatin pellet, and so by sequencing the 3′ end, he could map all RNAPII species, including those that are invisible to run-on methods because they are stalled and backtracked. Indeed, Chris' simple "3′NT" method allowed him to show that essentially all RNAPII stalls and backtracks when it encounters each nucleosome in its path (Weber et al. 2014). Analysis of 3′NT data also allowed Chris and postdoc Srinivas Ramachandran to infer the biological role of H2A.Z, namely to modulate the nucleosome barrier to transcription. Chris' 3′NT method precisely maps the addition of each base to the growing RNA chain within the active site of RNAPII without the need for antibodies or tags, making it potentially suitable for many applications. Meanwhile, Sheila Teves devised a high-resolution protocol for mapping DNA torsion genome-wide (Teves and Henikoff 2014), which, together with CATCH-IT and 3′NT, provides a powerful set of mapping tools for elucidating transcriptional dynamics in vivo.

To take advantage of the potential for massively parallel short-read sequencing, Jorja Henikoff, graduate student Kristina Krassovsky, and I developed a simple modification of MNase-seq with fragment midpoint-by-length display to achieve base-pair resolution of the full chromatin landscape (Henikoff et al. 2011; Henikoff and Henikoff 2012; Krassovsky et al. 2012). This native chromatin became input for high-resolution ChIP-seq of nucleosome remodelers by postdoc Gabe Zentner (Zentner et al. 2013) and for transcription factors by graduate student Siva Kasinathan (Kasinathan et al. 2014).

Meanwhile, postdoc Pete Skene developed a high-resolution cross-linking ChIP-seq protocol to map highly insoluble complexes (Skene *et al.* 2014).

The ultimate mapping challenge is the centromere, as evidenced by the fact that centromeres remain as multi-megabase gaps in the current human genome assembly. My interest in centromeres and surrounding heterochromatin began with a suggestion from my postdoc mentor, Charles Laird, to look into the classical phenomenon of position-effect variegation, and our later interest in histone variant evolution led us to explore this final frontier of the genome. The chromatin foundations of centromeres are nucleosomes containing the cenH3 (CENP-A in mammals) histone variant, and beginning in 1999 Paul Talbert and others in the lab began to identify and characterize cenH3s from worms (Buchwitz *et al.* 1999), flies (Henikoff *et al.* 2000), *Arabidopsis* (Talbert *et al.* 2002), and rice (Nagaki *et al.* 2003). I was intrigued by the poor sequence conservation of cenH3s despite the near perfect conservation of canonical H3 and variant H3.3. Postdoc Harmit Malik discovered that the basis for divergence is positive selection, which was unprecedented for a protein that is essential for every cell division (Malik and Henikoff 2001) and led to the concept of "centromere drive" (Malik and Henikoff 2009). In his own lab at the Fred Hutchinson Cancer Research Center, Harmit Malik has gone on to test predictions of the centromere drive hypothesis, including possible mechanisms for centromere evolution and the origin of species (Bayes and Malik 2009), while efforts in my lab have focused on characterizing cenH3 nucleosomes.

Studies by postdocs Yamini Dalal and Takehito Furuyama led to a model for the cenH3 particle as a right-handed half-nucleosome ("hemisome") (Dalal *et al.* 2007; Furuyama and Henikoff 2009), which we have confirmed in budding yeast using high-resolution native ChIP (Krassovsky *et al.* 2012), *in vitro* reconstitution (Furuyama *et al.* 2013), and *in vivo* chemical cleavage mapping (Henikoff *et al.* 2014). Meanwhile, postdoc Florian Steiner's mapping of worm holocentromeres revealed them to be dispersed cenH3 single-wrap particles, each resembling a yeast "point" centromere (Steiner and Henikoff 2014), which suggests that worm holocentromeres are polycentric, an idea first proposed by Franz Schrader 80 years ago (Schrader 1935). Surprisingly, insect holocentromeres entirely lack cenH3, as shown by postdoc Anna Drinnenberg, who is jointly mentored with Harmit Malik (Drinnenberg *et al.* 2014).

Our recent progress in applying high-resolution mapping tools to centromeres of yeast, rice, and worms has encouraged us to analyze ChIP-seq data by *de novo* computational mapping to explore the tandemly repetitive α-satellite sequences in which human centromeres are embedded—our genome's "black hole" (Henikoff 2002). Here we encountered another surprise: precisely phased single-wrap CENP-A particles occupying young α-satellite dimeric repeat units, in contrast to the diffusely phased CENP-A particles occupying older pericentric higher-order repeats (Henikoff *et al.* 2015). Our finding that two dimeric units dominate functional human centromeres and show decreasing abundance with evolutionary age provides a fascinating glimpse into the dynamic genetic process that continuously generates α-satellite. I am optimistic that we will soon be able to connect DNA sequence to centromere function in this final frontier of the genetic map.

Literature Cited

Ahmad, K., and S. Henikoff, 2002   The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell 9: 1191–1200.

Bayes, J. J., and H. S. Malik, 2009   Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science 326: 1538–1541.

Buchwitz, B. J., K. Ahmad, L. L. Moore, M. B. Roth, and S. Henikoff, 1999   A histone-H3-like protein in C. elegans. Nature 401: 547–548.

Conerly, M. L., S. S. Teves, D. Diolaiti, M. Ulrich, R. N. Eisenman *et al.*, 2010   Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome Res. 20: 1383–1390.

Cooper, J. L., E. A. Greene, B. J. Till, C. A. Codomo, B. T. Wakimoto *et al.*, 2008   Retention of induced mutations in a *Drosophila* reverse-genetic resource. Genetics 180: 661–667.

Dalal, Y., H. Wang, S. Lindsay, and S. Henikoff, 2007   Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol. 5: e218.

Deal, R. B., and S. Henikoff, 2010   A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev. Cell 18: 1030–1040.

Deal, R. B., J. G. Henikoff, and S. Henikoff, 2010   Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science 328: 1161–1164.

Drinnenberg, I. A., D. deYoung, S. Henikoff, and H. S. Malik, 2014   Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife 3: e03676.

Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind *et al.*, 2010   Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143: 212–224.

Furuyama, T., and S. Henikoff, 2009   Centromeric nucleosomes induce positive DNA supercoils. Cell 138: 104–113.

Furuyama, T., C. A. Codomo, and S. Henikoff, 2013   Reconstitution of hemisomes on budding yeast centromeric DNA. Nucleic Acids Res. 41: 5769–5783.

Gehring, M., K. Bubb, and S. Henikoff, 2009   Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447–1451.

Henikoff, J. G., J. A. Belsky, K. Krassovsky, D. M. Macalpine, and S. Henikoff, 2011   Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. USA 108: 18318–18323.

Henikoff, J. G., J. Thakur, S. Kasinathan, and S. Henikoff, 2015   A unique chromatin complex occupies young α-satellite arrays of human centromeres. Sci. Adv. 1: e1400234.

Henikoff, S., 2002   Near the edge of a chromosome's "black hole." Trends Genet. 18: 165–167.

Henikoff, S., and L. Comai, 1998   A DNA methyltransferase homolog with a chromodomain exists in multiple polymeric forms in *Arabidopsis*. Genetics 149: 307–318.

Henikoff, S., and J. G. Henikoff, 1992   Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89: 10915–10919.

Henikoff, S., and J. G. Henikoff, 2012   'Point' centromeres of *Saccharomyces* harbor single centromere-specific nucleosomes. Genetics 190: 1575–1577.

Henikoff, S., M. A. Keene, K. Fechtel, and J. W. Fristrom, 1986 Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell 44: 33–42.

Henikoff, S., K. Ahmad, J. S. Platero, and B. van Steensel, 2000 Heterochromatic deposition of centromeric histone H3-like proteins. Proc. Natl. Acad. Sci. USA 97: 716–721.

Henikoff, S., J. G. Henikoff, A. Sakai, G. B. Loeb, and K. Ahmad, 2009 Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res. 19: 460–469.

Henikoff, S., S. Ramachandran, K. Krassovsky, T. D. Bryson, C. A. Codomo et al., 2014 The budding yeast centromere DNA element II wraps a stable Cse4 hemisome in either orientation in vivo. Elife 3: e01861.

Henry, G. L., F. P. Davis, S. Picard, and S. R. Eddy, 2012 Cell type-specific genomics of Drosophila neurons. Nucleic Acids Res. 40: 9691–9704.

Kasinathan, S., G. A. Orsi, G. E. Zentner, K. Ahmad, and S. Henikoff, 2014 High-resolution mapping of transcription factor binding sites on native chromatin. Nat. Methods 11: 203–209.

Krassovsky, K., J. G. Henikoff, and S. Henikoff, 2012 Tripartite organization of centromeric chromatin in budding yeast. Proc. Natl. Acad. Sci. USA 109: 243–248.

Malik, H. S., and S. Henikoff, 2001 Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157: 1293–1298.

Malik, H. S., and S. Henikoff, 2009 Major evolutionary transitions in centromere complexity. Cell 138: 1067–1082.

McCallum, C. M., L. Comai, E. A. Greene, and S. Henikoff, 2000 Targeted screening for induced mutations. Nat. Biotechnol. 18: 455–457.

Mendel, G., 1866 Experiments in plant hybridization. Reprinted in 1951. J. Hered. 42: 3–47 (in German).

Mito, Y., J. Henikoff, and S. Henikoff, 2005 Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. 37: 1090–1097.

Mito, Y., J. Henikoff, and S. Henikoff, 2007 Histone replacement marks the boundaries of cis-regulatory domains. Science 315: 1408–1411.

Nagaki, K., P. B. Talbert, C. X. Zhong, R. K. Dawe, S. Henikoff et al., 2003 Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics 163: 1221–1225.

Ng, P. C., and S. Henikoff, 2001 Predicting deleterious amino acid substitutions. Genome Res. 11: 863–874.

Schrader, F., 1935 Notes on the mitotic behavior of long chromosomes. Cytologia (Tokyo) 6: 422–430.

Skene, P. J., A. E. Hernandez, M. Groudine, and S. Henikoff, 2014 The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife 3: e02042.

Steiner, F. A., and S. Henikoff, 2014 Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. eLife 3: e02025.

Steiner, F. A., P. B. Talbert, S. Kasinathan, R. B. Deal, and S. Henikoff, 2012 Cell type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Res. 22: 766–777.

Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J. Exp. Zool. 14: 43–59.

Talbert, P. B., R. Masuelli, A. P. Tyagi, L. Comai, and S. Henikoff, 2002 Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. Plant Cell 14: 1053–1066.

Teves, S. S., and S. Henikoff, 2011 Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev. 25: 2387–2397.

Teves, S. S., and S. Henikoff, 2014 Transcription-generated torsional stress destabilizes nucleosomes. Nat. Struct. Mol. Biol. 21: 88–94.

Till, B. J., S. H. Reynolds, E. A. Greene, C. A. Codomo, L. C. Enns et al., 2003 Large-scale discovery of induced point mutations with high throughput TILLING. Genome Res. 13: 524–530.

Till, B. J., S. H. Reynolds, C. Weil, N. Springer, C. Burtner et al., 2004 Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol. 4: 12.

Tran, R., J. Henikoff, D. Zilberman, R. Ditt, S. Jacobsen et al., 2005 DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr. Biol. 15: 154–159.

Tsai, H., T. Howell, R. Nitcher, V. Missirian, B. Watson et al., 2011 Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol. 156: 1257–1268.

van Steensel, B., and S. Henikoff, 2000 Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. Nat. Biotechnol. 18: 424–428.

van Steensel, B., J. Delrow, and S. Henikoff, 2001 Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. 27: 304–308.

Weber, C. M., J. G. Henikoff, and S. Henikoff, 2010 H2A.Z nucleosomes enriched over active genes are homotypic. Nat. Struct. Mol. Biol. 17: 1500–1507.

Weber, C. M., S. Ramachandran, and S. Henikoff, 2014 Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol. Cell 53: 819–830.

Yang, F., C. J. Kemp, and S. Henikoff, 2013 Doxorubicin enhances nucleosome turnover around promoters. Curr. Biol. 23: 782–787.

Yang, F., C. J. Kemp, and S. Henikoff, 2015 Anthracyclines induce double-strand DNA breaks at active gene promoters. Mutat. Res. 773: 9–15.

Zemach, A., I. E. McDaniel, P. Silva, and D. Zilberman, 2010 Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919.

Zentner, G. E., T. Tsukiyama, and S. Henikoff, 2013 ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. PLoS Genet. 9: e1003317.

Zilberman, D., D. Coleman-Derr, T. Ballinger, and S. Henikoff, 2008 Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456: 125–129.