Oligopaints: highly programmable oligonucleotide probes for visualizing genomes in situ

A dissertation presented

by

Brian Joseph Beliveau

to

the Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University
Cambridge, Massachusetts
December 2014
Oligopaints: highly programmable oligonucleotide probes for visualizing genomes in situ

Abstract

Fluorescence in situ hybridization (FISH) is a powerful assay that can visualize the position of DNA and RNA molecules in individual cells. Here, I describe the development of a method that utilizes complex oligonucleotide (oligo) libraries as a renewable source of FISH probes, which we term ‘Oligopaints’. Our novel FISH platform includes a reliable and robust protocol for the bulk production of fluorescently labeled, strand-specific, single-stranded DNA (ssDNA) probe sets and a bioinformatic pipeline able to identify optimal target sequences for in situ hybridization on a genome-wide scale. A key advantage of Oligopaints is that it permits the researcher to precisely define the genomic sequence contained within each probe molecule, specify the placement of fluorophores, and engineer ssDNA overhangs to which activities can be targeted. We harness this control to make two significant technological advances in FISH-based imaging. In one, Oligopaint probes are programmed to carry 5’ ssDNA overhangs that enable stochastic super-resolution microscopy via two methodologies, STORM and DNA-PAINT. We have used these probes to produce <25 nm resolution images of developmentally regulated chromatin in Drosophila and mouse, which are to our knowledge the first images at this resolution of single-copy chromosomal regions produced by FISH. In the second, we utilize single nucleotide polymorphism (SNP) data to generate FISH probes that can for the first time
visually distinguish single-copy regions of the maternal and paternal homologous chromosomes, thus allowing the examination of parent-of-origin dependent effects on chromosome positioning and gene expression in individual cells.
Acknowledgements

I have been very lucky to benefit from the patient and dedicated mentorship of Chao-ting Wu, who has taught me countless lessons about how to design experiments, interpret data, and present results in both written and interactive formats. Throughout my time, Ting has always challenged me to “think big,” which has been immensely helpful for my development. Above all, Ting has provided wonderful example of how to be an extremely rigorous scientist while never ceasing to be open-minded.

The Wu laboratory was a fantastic environment to work in. I am very grateful to have overlapped with some really great colleagues over my time, including Anna, Kenny, Matt, Sonny, Eric, Feyza, Ruth, Niroshi, Chamith, Nick, Fred, Caroline, Mohammed, Hien, Jelena, and Kwasi.

I had a great time meeting with my dissertation advisory committee: Steve Elledge, Bob Kingston, and Danesh Moazed. I am very grateful for the time my committee took to offer me excellent feedback and advice.

I would also like to thank my defense committee - Danesh Moazed, Giovanni Bosco, Suzanne Gaudet, and Jon Seidman – for taking the time to read this dissertation and conduct my examination.

Finally, I would like to thank my family for all the love and support. I am very lucky to have Ruth, Ali, Don, and Eva in my life.
# Table of Contents

## Chapter 1: Introduction

- Chromosomes as visible structures .......................................................... 2
- The first reports of in situ hybridization .................................................. 3
- Limitations of the early ISH methods ....................................................... 5
- ISH protocol refinement and the advent of recombinant DNA technology .................................................. 6
- Non-isotropic in situ hybridization with a fluorescent label ...................... 7
- The first report of FISH using a direct label ............................................ 9
- The introduction of hapten-labeled probes .............................................. 9
- Highly sensitive FISH via indirect labeling with signal amplification ......... 11
- Further refinement of the FISH protocol ............................................... 12
- Three-dimensional FISH ........................................................................ 13
- Early uses of oligos as ISH probes ....................................................... 15
- Technical barriers limiting the adoption of oligo-based probes for DNA FISH .................................................. 16
- The development of massively parallel oligo synthesis technology ........ 17
- The microarray revolution ...................................................................... 19
- From chip to solution: applications of pooled oligo libraries .................. 20
- Potential advantages of oligo-based DNA FISH probes ......................... 21
- FISH with oligo pools ................................................................. 21
- Advances in microscopy ......................................................................... 22
- Super-Resolution microscopy .................................................................. 24
- A new platform for oligo-based FISH ................................................... 25

## Chapter 2: Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes .............. 26
Chapter 4: Conclusion

Continuing to expand the utility of Oligopaints

Targeting smaller genomic intervals with Oligopaints

Nanoscale imaging of chromosomes with SMLM and Oligopaints

Quantitative imaging of chromosome structure with SMLM and Oligopaints

Moving beyond FISH

References

Appendix 1: Supplementary material related to Chapter 2

Appendix 2: Supplementary material related to Chapter 3

Appendix 3: Additional graduate school publications
Chapter 1:

Introduction
Chromosomes as visible structures

The field of chromosome biology has long been intimately connected to visual techniques. Indeed, the terms ‘chromatin’ (Flemming, 1879) and ‘chromosome’ (Waldeyer, 1888) originate from the gross morphological appearance of stained chromosomal DNA in histological preparations of interphase and metaphase cells, respectively (Flemming, 1879; Flemming, 1882; Waldeyer, 1888; reviewed in Paweletz, 2001). Moreover, the pioneering work of early cytologists allowed the role of the chromosome in inheritance to be postulated (Sutton, 1902; Sutton, 1903; Boveri, 1904) and demonstrated (Morgan et al., 1915) well before the chromosome was shown to contain DNA (Feulgen and Rossenbeck, 1924) and decades before the molecular structure of double-stranded DNA was first described (Watson and Crick, 1953). Subsequent advances in cell culture (Nowell, 1962; Nowell and Hungerford, 1962; Moorhead et al., 1962) and in the cytological preparation of metaphase chromosomes (reviewed in Hsu, 1979) set the stage for the development of a family of techniques that use nucleic acid hybridization to selectively visualize portions of chromosomes in fixed samples in a sequence-specific manner. These in situ hybridization techniques have made enormous contributions to fields such as gene mapping, inheritance, cytogenetics, genome stability, and nuclear organization. The origin, refinement, and broader impact of these techniques has been chronicled and reviewed extensively (Gall and Pardue, 1971; Jones, 1973; Coghlan et al., 1985; van der Ploeg and Raap, 1988; Lichter and Ward, 1990; Lichter et al., 1991; McNeil et al., 1991; Trask, 1991; van der Ploeg, 2000; Trask, 2002; Levsky and Singer, 2003; Volpi and Bridger, 2008; Riegel, 2014); thus, this introduction will focus more narrowly on technical advances in the design, synthesis, and hybridization of nucleic acid probes as well as developments in related areas such as improvements in the chemical synthesis of nucleic acids and fluorescent microscopy.
The first reports of in situ hybridization

The in situ hybridization (ISH) technique was first reported by Gall and Pardue in June of 1969 (Gall and Pardue, 1969) (Figure 1.1A); this report was quickly followed by publications from John, Birnsteil, and Jones in August of 1969 (John et al., 1969) (Figure 1.1B), a second publication from Gall and Pardue in October of 1969 (Pardue and Gall, 1969), and one from Buongiorno-Nardelli and Amaldi in March of 1970 (Buongiorno-Nardelli and Amaldi, 1970) (Figure 1.1C). These groups all used essentially the same approach to target ribosomal DNA (rDNA) in Xenopus oocytes (Gall and Pardue, 1969; John et al., 1969; Pardue and Gall, 1969), HeLa cells (John et al., 1969), and Chinese hamster tissue sections (Buongiorno-Nardelli and Amaldi, 1970): cells were cultured in the presence of $^{3}$H-labeled nucleotides, total RNA or DNA was extracted, rRNA or rDNA was isolated using isopycnic sedimentation in cesium chloride or sucrose gradients (Meselson et al., 1957), and then, finally, the radiolabeled rRNA or rDNA was added as a probe to fixed cytological sections whose chromosomal DNA had been denatured by alkaline treatment or heat. These proof-of-principle reports of ISH were followed by a quick burst of publications that demonstrated the utility of ISH by mapping the genomic location of satellite DNA sequences (Jones, 1970; Jones and Roberston, 1970; Pardue et al., 1970; Pardue and Gall, 1970; Henning et al., 1970; Gall and Pardue, 1971; reviewed in Jones, 1973) (Figure 1.1D,E).

The choice of ‘reiterated’ sequences such as the rDNA and satellites as ISH targets was ideal for several reasons. First, probe material could readily be isolated in a pure form by density gradient centrifugation. Additionally, these highly repetitive target sequences were known to have very fast reannealing kinetics in vitro following denaturation (Gillespie and Spiegelman, 1965; Britten and Kohne, 1968), thus allowing for hybridization to occur in situ over relatively
quick timescales (e.g. <1 day) (reviewed in Gall and Pardue, 1971; Jones, 1973). Finally, as the repeats chosen occur in large (~hundreds of kilobase to many megabases), contiguous segments in the genome, these early studies were able to generate images in which the

Figure 1.1 Early examples of in situ hybridization. (A,B) Amplified rDNA is visualized in Xenopus oocytes. (A) is from (Gall and Pardue, 1969); (B) is from (John et al., 1969). (C) rDNA is visualized in liver tissue sections from a Chinese hamster. Image is from (Buongiorno-Nardelli and Amaldi, 1970). (D,E) Mouse satellite DNA is shown to reside near the centromeres of the acrocentric mitotic mouse chromosomes. (D) is from (Pardue and Gall, 1970); (E) is from (Jones and Robertson, 1970). Note the clusters of radioactive signal overlaying the chromocenters in the interphase nuclei. Scale bar in (D) is 5 µm. All images are reproduced with permission.
majority of the radioactive signal was localized focally, making the data readily interpretable. The choice of mouse satellite DNA was particularly fortuitous, as these satellites were discovered to cluster at just a few sites in interphase nuclei despite being observed on every chromosome besides the Y in metaphase (Jones, 1970; Jones and Roberston, 1970; Pardue and Gall, 1970; reviewed in Jones, 1973) (Figure 1.1D,E). Indeed, early ISH studies provided some of the first insights into interphase nuclear organization and the occurrence of interchromosomal (Jones, 1970; Jones and Roberston, 1970; Pardue and Gall, 1970) and intrachromosomal (Pardue et al., 1970; Hennig et al., 1970) interactions amongst heterochromatic sequences that are still being actively studied today (reviewed in Politz et al., 2013).

**Limitations of the early ISH methods**

Although the early ISH experiments yielded important and prescient observations about the organization of repetitive sequences in the genome, technical limitations hindered the application of ISH to a broader range of targets. One limitation was the inability to easily isolate and amplify probe RNA or DNA specific for a non-repetitive target. This problem was compounded by the use of $^{3}$H as the means of labeling and detection: the relatively weak and infrequent radioactive decay of $^{3}$H and difficulties generating probes with high specific activities necessitated very long exposures, often on the order of weeks or even months (Gall and Pardue, 1969; John et al., 1969; Pardue and Gall, 1969; Buongiorno-Nardelli and Amaldi, 1970; reviewed in Gall and Pardue, 1971), which in turn made the detection of ‘non-reiterated’ genomic intervals such as individual genes difficult or impossible. Furthermore, the autoradiographic detection methods used to register the $^{3}$H signal offered a very poor spatial resolution, which dramatically limited the ability of the technique to map sequences onto highly
condensed metaphase chromosomes (reviewed in Coghlan et al., 1985; Trask, 2002). Taken together, these technical hurdles stagnated the development of the ISH field. Indeed, it took more than a decade for ISH to transition from highly repetitive targets to the first genomically unique targets, a ~5 kilobase portion of the β globin gene that was observed to map to chromosome 11 (Malcolm et al., 1981), and a ~15 kilobase segment of uncharacterized sequence that was observed to map to chromosome 1 (Harper and Saunders, 1981).

**ISH protocol refinement and the advent of recombinant DNA technology**

A number of key advances occurred during the 1970s that buoyed the improvement of ISH. Perhaps the most important development was the invention of recombinant DNA technology and molecular cloning. This technology allowed researchers to fragment genomic DNA with restriction endonucleases (Meselson and Yuan, 1968; Arber and Linn, 1969; Smith and Wilcox, 1970; Kelly and Smith, 1970; Danna and Nathans, 1971) and then insert the fragments into a vector such as a virus (Jackson et al., 1972), plasmid (Cohen et al., 1973), or cosmid (Collins and Hohn, 1978) by DNA ligation (Weiss and Richardson, 1967; Lobban and Kaiser, 1973). The fragment could then be amplified by cell or bacterial culture, thus enabling large-scale, relatively pure preparations of recombinant probe material to be produced.

A new method of incorporating radiolabeled nucleotides into the amplified probe, termed ‘nick translation,’ (Rigby et al., 1977) further enhanced probe production. In a nick translation reaction, the double-stranded DNA (dsDNA) to be labeled is first treated with DNAase I (Young and Sinsheimer, 1965) to introduce internal single-strand nicks. The nicked dsDNA is then incubated with radiolabeled nucleotides and *E. coli* DNA polymerase I (Kornberg 1969), which excises the base immediately downstream of the nick and repairs the lesion via the
incorporation of a radiolabeled nucleotide (Kelly et al., 1970). Additionally, as not all nicks are repaired under the conditions used, this method also fragments the probe into smaller pieces (~400 bases), which aids its diffusion into cells and tissues during ISH reactions (Rigby et al., 1977). Nick translation facilitated the generation of probes with specific activities two orders of magnitude or more greater than those produced by previous methods (Gelb et al. 1971; Rigby et al., 1977; reviewed in Coghlan et al., 1985).

As the ISH field matured, modifications to the original hybridization protocols emerged that improved the performance of the technique. One of the first innovations was the inclusion of the chemical denaturant formamide during the denaturation and hybridization steps (Jones and Robertson, 1970). Formamide was observed to lower the melting temperature ($T_M$) of RNA-DNA and DNA-DNA duplexes by 0.72°C per % formamide in solution (McConaughy et al., 1969) and thus provided a convenient and simple way to increase the effective temperature of a hybridization or wash by 30°C or more (Figure 1.2). Another very important innovation was the use of the molecular crowding agent dextran sulfate in the hybridization buffer, which was observed to increase the rate of nucleic acid renaturation by 10-100 fold in solution (Wetmur, 1975; Wahl et al., 1979).

**Non-isotropic in situ hybridization with a fluorescent label**

As discussed above, the use of $^3$H as the labeling reagent hindered the development of ISH. In the late 1970s and early 1980s, two approaches were taken to try and improve the speed and sensitivity of ISH. One strategy utilized radioisotopes with higher energy $\gamma$ emissions such as $^{125}$I (.035 MeV) (Prensky et al., 1973) or $\beta$ emissions such as $^{32}$P (1.7 MeV) (Hudson et al., 1981), as these were easier to detect than the weak $^3$H $\beta$ emissions (.018 MeV) (Randerath,
Of these, $^{32}\text{P}$ was particularly useful, as its relatively quick decay (half-life $\sim$2 weeks vs. $\sim$12 years for $^3\text{H}$), coupled with its higher energy emissions, allowed for shorter exposures on the order of days instead of weeks to months (reviewed in Coghaln et al., 1985). An alternative approach was the use of ‘non-isotropic’ labeling and detection methods. The first major breakthrough in this area came from Rudkin and Stollar (1977). Stollar had previously described the generation of rabbit antiserum raised against RNA-DNA hybrids (Stollar, 1970). In their experiment, Rudkin and Stollar prepared salivary poltyene chromosomes from Drosophila melanogaster – these chromosomes were fixed, treated with RNase to remove all endogenous RNA, and denatured with a combination of heat and formamide. They then added unlabeled 5S rRNA that had been purified from Drosophila.
embryos as the hybridization probe, followed by the addition of the antiserum specific for RNA-DNA hybrids. Finally, they visualized their hybridization experiment by the addition of a fluorescent rhodamine-conjugated secondary antibody (Figure 1.3A), in what is considered to be the earliest example of fluorescence in situ hybridization (FISH). Because the rhodamine label was conjugated to the secondary antibody instead of directly to the nucleic acid hybridization probe, this type of approach became to be known as ‘indirect’ labeling and would soon transform the ISH field.

The first report of FISH using a direct label

The next innovation to follow that of Rudkin and Stollar came from the laboratory of Pieter van Duijn (Bauman et al., 1980). Rather than using the indirect immunofluorescence method, van Duijn and colleagues took advantage of a recently developed chemical method for conjugating moieties to the 3’ end of RNA molecules (Hansske and Cramer, 1979) to attach a fluorescent tetramethyl rhodamine isothiocyanate (TRITC) molecule directly to purified RNA. This strategy was used to visualize kinetoplast DNA in the trypanosome Crithidia luciliae, viral DNA in infected human culture cells, and ribosomal DNA on Drosophila hydei salivary polytene chromosomes (Figure 1.3B) using directly labeled probes (Bauman et al., 1980). Although the ISH community did not quickly adapt the direct fluorescent labeling method, its concept would be instrumental to the later development of oligonucleotide-based hybridization probes.

The introduction of hapten-labeled probes

Shortly after the first report of van Duijn’s direct labeling method, a set of important studies was published by the laboratory of David Ward. The Ward laboratory had previously worked on the
Figure 1.3 Non-isotropic in situ hybridization. (A,B) 5S rRNA is visualized on Drosophila salivary polytene chromosomes using fluorescence microscopy. (A) is from (Rudkin and Stollard, 1977) and shows Drosophila melanogaster chromosomes. (B) is from (Bauman et al., 1980) and shows Drosophila hydei chromosomes. Fluorescent ISH micrographs are shown in the top images while the same field of view is shown with general chromosome stains (aceto-orecin and DAPI, respectively) below. (C) Mouse satellite DNA is visualized on mitotic chromosome spreads using fluorescence microscopy. Image is from (Manuelidis et al., 1982). (D) Fluorescent micrograph of a human metaphase spread in which the pericentromeric heterochromatin of chromosome 1 is visualized by the method of Pinkel and Gray (yellow). Chromosomes are stained with propidium iodide (red). Image is from (Waye et al., 1987). All images are reproduced with permission.
production of chemically modified nucleotide analogs such as those containing mercury (Dale et al., 1973) or sulfur (Livingston et al., 1976). In 1981, the Ward laboratory described the synthesis of the ribonucleotide UTP and the deoxyribonucleotide dUTP with covalently attached biotin molecules; they referred to the biotin molecule as a “hapten” in this context as the biotin is a small molecule that can be recognized immunogenically (Langer et al., 1981). Ward and colleagues further demonstrated that DNA and RNA polymerases could incorporate these “haptenized” nucleotides in vitro, and that the haptenized nucleotides were compatible with the nick translation method of probe labeling (Rigby et al., 1977; Langer et al., 1981). They subsequently used biotinylated probes generated by nick translation to visualize sequences on D. melanogaster polytene chromosomes (Langer-Safer et al., 1982), actin mRNA in chicken muscle tissue culture (Singer and Ward, 1982), and mouse satellite DNA (Manuelidis et al., 1982) (Figure 1.3C) using indirect immunofluorescence. While this method and the related direct labeling method of van Duijn had not yet improved the sensitivity of ISH, they importantly reduced the required exposure time from days or weeks to mere seconds and eliminated the need to use potentially dangerous radioisotopes.

**Highly sensitive FISH via indirect labeling with signal amplification**

Perhaps surprisingly, Ward and colleagues did not detect their biotinylated probes using the protein avidin, which was known to bind biotin with a remarkably high affinity for a non-covalent interaction ($K_d \approx 10^{-14}$ M) (Green, 1975); instead, the biotin was detected with an anti-biotin antibody that was raised as part of their initial study (Langer-Safer et al., 1982). Ward and colleagues mention that they had originally intended to use avidin, but observed very high levels of background staining with that approach (Langer-Safer et al., 1982). This background was likely related to an aspect of the protocol that they used, such as the lack of a blocking step prior
to the addition of avidin or the use of phosphate-buffered saline as the buffer for the avidin incubation (Singer et al., 1987). In any case, biotinylated probes were successfully visualized using fluorescent fluorescein-conjugated avidin by Pinkel and Gray (Pinkel et al., 1986a; Pinkel et al., 1986b, Waye et al., 1987) (Figure 1.3D). Moreover, Pinkel and Gray established that the fluorescent signal could be further amplified by detecting the fluorescein-avidin with the sequential addition of a biotinylated anti-avidin antibody, which introduced more binding sites for fluorescein-avidin; this process could be repeated for several rounds to ultimately recruit up to ~200 fluorescein molecules per detected biotin (Pinkel et al., 1986b). Using this technology, Pinkel and Gray were able to visualize genomically dispersed blocks of the human 28s rDNA repeat on metaphase chromosomes, some of which were estimated to be as small as ~20-30 kilobases (Pinkel et al., 1986b). This experiment provided the first demonstration that in situ hybridization detected by fluorescence microscopy could offer a sensitivity that approached that of autoradiographic detection methods, which could detect sequences on the kilobase scale (Malcolm et al., 1981).

**Further refinement of the FISH protocol**

Work from many laboratories in the 1980s and early 1990s, including large contributions from those of Mels van der Ploeg (reviewed in van der Ploeg and Raap, 1988), Robert Singer (reviewed in McNeil et al., 1991), and David Ward (reviewed in Lichter and Ward, 1990; Lichter et al., 1991), helped to transform FISH into a broadly used technique. The development of the protocol was particularly aided by an extensive series of optimizations performed by Jeanne Lawrence and Robert Singer (1985), in which the effects of the sample fixation method, pre-hybridization treatment, hybridization duration, probe concentration, wash conditions, mean probe fragment size, and the propensity of the probe to form probe ‘networks’ of interacting
labeled molecules at the site of hybridization (Wahl et al., 1979) were systemically examined. Lawrence and Singer found that 4% paraformaldehyde was optimal as the fixative, probe fragments <~1000 bp in size gave the lowest background (~20+:1 signal:noise), and biotinylated probes in particular were very sensitive to fragment size, with only fragments with a size of <~150 bases giving strong staining over background (10:1 signal:noise) (Lawrence and Singer, 1985). Another important innovation was the invention of “suppression” hybridization by the van der Ploeg laboratory (Landegent et al., 1987). As FISH was typically performed using nick-translated clones of genomic DNA, the probe fragments often contained homology to highly repetitive sequences in addition to the target locus or RNA, as up to half of complex mammalian genomes were estimated to be composed of interspersed repeats (Britten and Kohne, 1968). In order to avoid the spurious annealing of probe fragments to repetitive sequences, unlabeled repetitive DNA (e.g. the fraction with a C_dt value ≤1 [Britten and Kohne, 1968]) from the species of interest was added during the hybridization (Landegent et al., 1987; Lichter et al., 1988; Pinkel et al., 1988). This strategy greatly improved the reliability of clone-based probes irrespective of their repetitive DNA content and thus obviated the routine need for the laborious screening of clones for repetitive content (reviewed in McNeil et al., 1991).

**Three-dimensional FISH**

The rapid developments discussed above led to the creation and adoption of FISH protocols that are remarkably similar to many of the protocols still employed today (reviewed in Volpi and Bridger, 2008; Riegel, 2014). One important refinement of the FISH protocol that occurred slightly later came from the laboratory of Thomas Cremer, who performed one of the earliest studies of chromosome organization in interphase nuclei using FISH (Cremer et al., 1986). As the Cremer lab has focused on understanding the consequences of the three-dimensional...
organization of genomic sequences in the interphase nucleus, they were concerned about the potential for the treatments used during FISH to destroy or alter the structure of the nucleus and thus render their FISH results uninterpretable. Motivated by this concern, they performed a systematic examination of the effects of fixation, pre-hybridization treatment, denaturation, and washing on chromosome structure and gross nuclear morphology using light and electron microscopy (Solovei et al., 2002). Cremer and colleagues observed that the heat denaturation step in particular was responsible for the most dramatic changes in the morphology of stained chromatin observed by electron microscopy, including the loss of aggregate structures and the appearance of a filamentous network; thus, their “3D FISH” protocol was developed to minimize the temperature used during denaturation and the duration of the denaturation step (Solovei et al., 2002; Cremer et al., 2008).

3D FISH has become the standard for investigations of chromosome organization using FISH. In brief, the protocol is performed as follows: First, cells are fixed in 4% paraformaldehyde for 10 minutes. Next, the cells are permeabilized in by detergent treatment in 0.5% Triton X-100. The cells can be further permeabilized by a series of snap-freezes in liquid nitrogen. The cells are then incubated briefly in 0.1 N HCl to increase probe accessibility via the formation of nicks in the chromosomal DNA and denaturation of proteins. Finally, a hybridization buffer containing probe, formamide, and dextran sulfate is added, the sample is denatured for 2-3 minutes at 75°C, and hybridization is allowed to proceed for at least one day. This protocol has been used with a range of probe types, including clone-based probes, probes generated by sets of adjacent polymerase chain reaction (PCR) amplicons (Saiki et al., 1985; Mullis et al., 1986; Martinez et al., 1996; Lamb et al., 1997), and whole-chromosome paints generated by degenerate oligo-primed PCR of flow sorted chromosomes (Bolzer et al., 2005). Slightly modified versions of this protocol (Lanzuolo et al., 2007; Beliveau et al., 2012; Beliveau et al.,
2014) were also used for several of the experiments presented in Chapter 2 and nearly all of the experiments presented in Chapter 3; these altered protocols were finely tuned to be optimal for different sets of oligonucleotide (oligo) probes, which as discussed below, can be designed to have specific thermodynamic properties.

**Early uses of oligos as ISH probes**

All of the ISH and FISH probes discussed so far were produced from isolated genomic material; in the case of the very early ISH studies, genomic RNA or DNA was used directly as the probe material, while in later studies the genomic sequences were cloned and then subsequently amplified and labeled. In addition to these sources, researchers have also used RNA or DNA oligos produced by chemical synthesis (Michelson and Todd, 1955) as probes. The first reports of oligo probes being used for radioactive ISH were enabled by the development of the phosphoramidite method of oligo synthesis in the laboratory of Marvin Caruthers (Matteucci and Caruthers, 1981; Beaucage and Caruthers, 1981; McBride and Caruthers, 1983; reviewed in Caruthers, 1985; Caruthers, 2011); unlike the reagents used in previous oligo synthesis methods, those used for phosphoramidite synthesis were chemically stable at room temperature, not prone to oxidation by air, and resistant to spontaneous hydrolysis under normal laboratory conditions (Beaucage and Caruthers, 1981). Accordingly, the phosphoramidite method soon became the standard oligo synthesis chemistry and was instrumental in the development of automated oligo synthesis systems, as the lability of the reagents used in previous methods demanded that each synthesis be conducted manually by an experienced chemist (reviewed in Reese, 2002; Caruthers, 2011).
The use of oligos synthesized by the phosphoramidite method in radioactive ISH was pioneered by J.P. Coghlan and colleagues, who used $^{32}$P labeled DNA oligos to identify cells expressing kallikrein mRNA in mouse kidney sections, insulin mRNA in rat kidney sections, calcitonin mRNA in rat thyroid tissue, and arginine vasopression mRNA in rat brain sections (Coghlan et al., 1985). This work was soon followed by a set of studies from several groups that used synthetic oligos labeled with biotin to visualize cells expressing arginine vasopression mRNA in the rat hypothalmus (Arai et al., 1988), proopiomelanocortin mRNA in mouse and rat pituitary sections (Larsson et al., 1988), mammalian telomeres in a range of species (Moyzis et al., 1988; Meyne et al., 1989), and human papilloma virus DNA in primary human tissue samples (Cubie and Norval, 1988) using indirect immunofluorescence or immunohistochemistry.

**Technical barriers limiting the adoption of oligo-based probes for DNA FISH**

Despite the success of the aforementioned studies that used oligo probes for ISH and FISH, several technical barriers prevented synthetic oligos from becoming a routine source of probes for DNA FISH experiments. One of the largest hurdles was general lack of genome-scale sequence databases, as knowledge of the target sequence was required *a priori* for the synthesis of a complementary probe molecule. Another issue was the throughput with which oligos could be efficiently synthesized, as oligos were typically produced individually using solid-phase synthesis (reviewed in Caruthers, 1985; Hougaard et al., 1997). Although individual oligos were shown to be sufficient to identify cells expressing marker mRNAs present in many (~100-1000+) copies (Arai et al., 1988; Larsson et al., 1988; Dirks et al., 1990; Dirks et al., 1991) and repetitive arrays of DNA (Moyzis et al., 1988; Meyne et al., 1989; Dernburg et al., 1996; Lansdorp et al., 1996), an individual oligo would not be predicted to have the sensitivity to visualize genomically unique regions of chromosomal DNA. Indeed, as some of the most
sensitive clone-based FISH probes targeted regions spanning on the order of ten kilobases using fragments with a median size of a few hundred bases (Pinkel et al., 1986b), dozens if not hundreds or more oligos would be required to produce a similar level of genomic coverage. Furthermore, these oligos would then need to be labeled with at least one fluorophore or hapten to reproduce the labeling density of clone-based probes, which would increase the potential cost and complexity of synthesis. Thus, performing DNA FISH using oligo probes against a genomically unique region was not technically tractable in the absence of a cost-effective way to produce and label a large number of oligos.

**The development of massively parallel oligo synthesis technology**

Automated oligo synthesis technology dates back to the early 1980s (Alvarado-Urbina et al., 1981; Hunkapiller et al., 1984; reviewed in Caruthers, 1985); these early machines could synthesize up to a handful of oligos using solid-phase phosphoramidite chemistry. While the throughput of this type of automated synthesis technology was increased during 1990s with the introduction of machines that could synthesize up to 96 oligos in parallel (Sindelar and Jaklevic, 1995; Lashkari et al., 1995; Rayner et al., 1998), progress in further expanding the number of oligos that could be synthesized in parallel stagnated due to technical difficulties in miniaturizing aspects of the synthesis reaction (Cheng et al., 2002). Fortunately, scientists at Affymetrix introduced a very exciting adaptation of the solid-phase synthesis process. Rather than trying to miniaturize solid-phase oligo synthesis columns, Stephen Fodor and colleagues developed a synthesis platform that relied on the concept of 'spatially addressable photolithography' prevalent in manufacture of high-density integrated circuitry such as that found in computer processors (Fodor et al., 1991).
In the Affymetrix method, chemical synthesis occurs in a set of positions arrayed on a solid surface such as a microscope slide. As in the Caruthers solid-state phosphoramidite method, oligos are polymerized one base at a time in the 3’ to 5’ direction by the addition of “protected” nucleotides; the “protecting” group prevents the further extension of the nascent oligo until it is chemically removed (reviewed in Caruthers, 1985). In the Affymetrix method, these protecting groups are photo-labile; thus, light can be directed via a photolithographic mask to specific positions on the surface in order to selectively deprotect only the nascent oligos arrayed in those positions while leaving the remainder of spots protected. The deprotected nascent oligos can then undergo one round of extension, and the process can then be repeated with a distinct photolithographic mask to direct the subsequent locations to be deprotected (Figure 1.4). In this way, many distinct oligos can be synthesized in parallel on the same solid substrate.

Figure 1.4 The Affymetrix ‘spatially addressable photolithography’ method for massively parallel oligo synthesis. (A) A schematic illustrating how a photolithographic mask can be used to selectively illuminate a subset of arrayed positions on a surface. (B) An illustration of how selective deprotection allows for multiple different oligos to be synthesized in parallel on the same surface. Figure is adapted from (Lipshutz et al., 1999) with permission.
The microarray revolution

Although the first published report of the Affymetrix array-based synthesis technology was primarily focused on peptide synthesis, it also included a proof-of-principle demonstration in which 13 dinucleotides were synthesized in parallel (Fodor et al., 1991). This report was quickly followed by reports of more complex oligo “microarrays” on which 256 distinct 8mers were generated (Fodor et al., 1993; Pease et al., 1994). Remarkably, just two years later, Affymetrix produced microarrays containing >135,000 15mers (Chee et al., 1996). This rapid progress, along with parallel contributions from the laboratories of Patrick Brown and Ronald Davis (Schena et al., 1995; Lashkari et al., 1997), helped to launch DNA microarray technology as a powerful tool for the study of gene expression and genetic variation (reviewed in Brown and Botstein, 1999; Gresham, 2008; Katsanis S and Katsanis N, 2013). Further refinement of photolithographic synthesis method soon pushed the technology even further, resulting in microarrays with as many as one million unique 25mer oligos (reviewed in Lipshutz et al., 1999). In addition to the Affymetrix method, several alternative array-based synthesis platforms emerged (reviewed in Kosuri and Church, 2014), including the maskless photolithography method of NimbleGen and LC Sciences (Singh-Gasson et al., 1999; Gao et al., 2001), an ink-jet printing method developed by Agilent (Blanchard et al., 1996; Hughes et al., 2001), and a semiconductor-based method created by CustomArray (Ghindilis et al., 2007). Today, these methods can be used to produce anywhere from a few thousand up to ~1 million unique oligos ranging in length from ~20 bases to nearly 200 bases (reviewed in Kosuri and Church, 2014). Excitingly, these oligos can be purchased commercially at a cost of <$0.001/base, which is 1-2 order of magnitudes cheaper per base than oligos produced by commercial column-based synthesis (Figure 1.5).
From chip to solution: applications of pooled oligo libraries

The ability to design and commercially order thousands to hundreds of thousands of unique oligos has inspired researchers to devise new applications for array-synthesized oligos. Many of these applications utilize the oligos not as a part of a traditional slide-based microarray, but instead as a mixed pool or 'library' of oligos in solution. Oligo pools have been used as building blocks for the assembly of synthetic genes (Tian et al., 2004; Zhou et al., 2004; Richmond et al., 2004; Kosuri et al., 2010; reviewed in Kosuri and Church, 2014), target capture probes for next generation sequencing (Porreca et al., 2007; Gnirke et al., 2009), components of DNA origami structures (Rothemund, 2006; Douglas et al., 2009; Ke et al., 2012), and high density information storage substrates (Church et al., 2012; Goldman et al., 2013). Of particular interest

Figure 1.5 The cost-savings offered by highly complex oligo pools compared to other modes of oligo synthesis. Over the length scales used for FISH (~25-200 bases), array-based oligo pools are orders of magnitudes cheaper than column-synthesized oligos. Figure is reproduced with permission from (Kosuri and Church, 2014).
to the work presented hereafter, oligo pools were also considered as a source of DNA FISH probe material by the Chao-ting Wu laboratory (Wu et al., 2010). This idea, which precedes the author’s time in the Wu laboratory, is the foundation upon which all of the work presented in Chapters 2 and 3 rests.

**Potential advantages of oligo-based DNA FISH probes**

The properties of oligo-based probes offer several potential advantages over more traditional clone- or PCR-based probes. One key advantage is the full control researchers have over the sequence of each synthetic molecule. This control can be harnessed to design probes with very specific thermodynamic properties (Santalucia, 1998) that are predicted to respond uniformly to changes in the hybridization and wash conditions. Furthermore, probes can be designed to lack homology to repetitive elements, thus eliminating the need for suppressive hybridization techniques. Additionally, as the probe source is entirely synthetic, there is no need to isolate and screen genomic clones or vet genomic PCR amplicons in order to target a region of interest; rather, only a working knowledge of the genomic sequence to be targeted is needed. Finally, due to their small size and inherent single-stranded nature, oligo probes have been suggested to diffuse into cells and tissues very efficiently (Dirks et al., 1990) and are not expected to form the ‘networked’ hybridization structures sometimes observed with clone-based FISH probes (Wahl et al., 1979; Lawrence and Singer, 1985).

**FISH with oligo pools**

At the onset of the work described in Chapters 2 and 3, the most prominent example of FISH performed with oligo pools was the single-molecule FISH method developed by Raj, van
Oudenaarden, and Tyagi (Raj et al., 2008). In their approach, Raj and colleagues used ~50 20mer oligos, each carrying one fluorophore, to visualize individual RNA molecules; this strategy was an adaption of a single-molecule RNA FISH method developed in the laboratory of Robert Singer that used five 50mer oligos carrying five fluorophores each per RNA molecule (Femino et al., 1998) (Figure 1.6A,B). Subsequent adaptations of this method have been used to visualize coordinated transcription across chromosomal domains (Levesque et al., 2013), perform high-throughput imaging of transcriptional changes in response to inflammatory signaling (Lee et al., 2014), and quantify allele-specific mRNA expression (Hansen and van Oudenaarden, 2013; Levesque et al., 2013). Additionally, these very exciting methods also provided a very valuable estimate of the number of oligos and fluorophores required to see a focal signal over background in fixed cells and tissues. The use of oligo pools for FISH was not limited to RNA targets, as two reports were published while we were performing the work presented Chapter 2 that demonstrated the use of oligo pools as probes for DNA FISH. The first came from researchers at Agilent, who employed pools of an unknown complexity of very long (>150mer) oligos to label chromosomal regions in human cells using a proprietary protocol (Yamada et al., 2011) (Figure 1.6C). In the second study, which was a collaboration between the laboratory of Wendy Bickmore and Roche NimbleGen, libraries of ~60-100 base DNA oligos designed for targeted sequence capture were end-labeled with a fluorescent dye and used for DNA FISH (Boyle et al., 2011) (Figure 1.6D). Collectively, these studies encouraged our efforts to develop our platform for oligo-based FISH.

**Advances in microscopy**

It is important to mention that the results described in Chapter 2 and 3 were enabled not only by the availability of inexpensive, high complexity oligo pools, but also by many advances in...
Figure 1.6 Examples of FISH with pooled oligo probes. (A,B) Single-molecule RNA FISH performed with oligo probes. (A) is adapted from (Femino et al., 1999) and uses 5 52-53mer oligos labeled 5 fluorophores each to visualize β-actin mRNA. The red arrow marks a fluorescent bead, while the yellow arrow denotes a bright nuclear focus that authors believe represent nascent transcription at the β-actin gene body. (B) is adapted from (Raj et al., 2011) and uses 48 singly-labeled oligos to visualize FKBP5 mRNA. A black box is placed over the original panel label for simplicity of presentation. Scale bar is 5 µm. (C) Two-color DNA FISH performed with long (>150mer) oligos targeting 23 kb (green) and 56 kb (red) of a tandem duplication in human cells. Image is from (Yamada et al., 2011). (D) Chromosome painting with a conventional chromosome paint targeting chromosome 2 (red) and a custom oligo pool targeting the exome of chromosome 2 (green) in mouse embryonic stem cells. Image is from (Boyle et al., 2011). Scale bar is 5 µm. All images are reproduced with permission.

fluorescence microscopy, image registration, and image processing. Indeed, many significant technical innovations have been introduced since the FISH protocol optimizations of Singer and Lawrence (1985) and Pinkel and Gray (Pinkel et al., 1986b) were performed. One key invention was the cooled, charge-coupled device (CCD) (Hiroka et al., 1987), which made it possible to efficiently collect quantitative information about photon emissions in a widefield focal plane over a broad spectrum of wavelengths. Another important development was laser scanning confocal microscopy (Cremer C and Cremer T, 1978; Brakenhoff et al., 1979; Shotton, 1989), which allowed researchers to examine samples in a small, scanning focal area with very little competing out-of-focus light. Laser scanning confocal microscopes have been further enhanced
by the development of more sophisticated and sensitive photomultiplier tube (PMT) detectors that convert photons into electrical signals more efficiently (~25-45% quantum efficiency) over a broader spectrum of wavelengths (~350-675 nm) (reviewed in Art, 2006).

Super-Resolution microscopy

While the work presented in Chapter 2 and 3 details some of our efforts to enhance the sensitivity and resolution of FISH, our ability to visualize signals by fluorescent microscopy is ultimately limited by diffraction to a focal spot of a minimum size of ~200 nm in the X-Y plane and ~500 nm in the Z direction. Thus, the volume of a diffraction-limited signal is considerably larger than many nuclear structures of interest to researchers. Given this limitation, assays designed to assess phenomena such as changes in chromatin structure using fluorescent microscopy have been largely restricted to measuring changes in the distance between diffraction-limited FISH signals separated by tens or hundreds of kilobases of genomic sequence (Lanzuolo et al., 2007; Eskeland et al., 2010) and thus can provide only minimal information about fine-scale structural differences. However, a number of techniques have been developed that allow ‘super-resolution’ imaging of features smaller than the diffraction limit, with some techniques able to resolve distances on the order of tens of nanometers (reviewed in Huang et al., 2009; Godin et al., 2014).

One family of super-resolution imaging techniques relies on the detection of stochastically occurring single-molecule fluorescence events, which are used to localize the spatial position of each fluorophore molecule with high precision. These single-molecule localization microscopy (SMLM) techniques have enormous potential to visualize nanoscale structural features in the nucleus, as their resolution is not limited by diffraction; instead, it is dependent on the amount of
light detected from each single-molecule fluorescence event, which determines the precision of the localization (Patwardhan, 1997; Thompson et al., 2002), and the labeling density of fluorophores on the structure to be imaged, which determines the maximum size of the spatial features that can be reliably determined according to the Nyquist criterion (the labeling density needs to be >2-fold higher than the feature size to avoid aliasing, i.e. the detection of artifactual structural features due to undersampling) (reviewed in Huang et al., 2009).

**A new platform for oligo-based FISH**

Here, I will describe our efforts to design, validate, and optimize computational and molecular biological pipelines for the creation of highly efficient, ssDNA FISH probes from complex oligo libraries. We then further show that the programmability of these oligos afforded by their synthetic nature has allowed us to push DNA FISH into new territories, including visualization of fine-scale chromatin structure using SMLM and the visual distinction of homologous chromosomes.
Chapter 2:

Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes
Chapter 2 Contributions

The work presented in this chapter is the result of a very nice collaboration with many other members of the C.-ting Wu laboratory. I am grateful for assistance my co-first author, Eric Joyce, who performed the FISH shown in Figure 2.3 panels E, F, in Figure 2.4 panel E, Figure 5 panels B-D, and Supplementary Figure A1.11. Eric also helped to develop the probe generation and FISH protocols, and helped to write the paper. This study was also greatly aided by Nicholas Apostolopoulos, who perform the FISH shown in Figure 2.2, Figure 2.4 panel A, and Supplementary Figure A1.9; Nick also help to analyze data and optimize the probe generation and FISH protocols. We also received invaluable assistance from Feyza Yilmaz, Chamith Fonska, and Ruth McCole, who helped to develop our genome-scale probe mining pipeline and who collectively wrote nearly all of the python scripts used to handle our probe files. I was also aided by some preliminary work performed by my co-authors Kenny Chang and Jin Billy Li that helped inform the approach we decided to take. This study would of course not be possible without the mentorship and guidance of C.-ting Wu, who conceived of the Oligopaints concept, supervised the project, provided economic support, created Supplementary Figure A1.12, and co-wrote the paper. As the primary driver of this project, I created the bioinformatic probe design pipeline, the probe production protocol, and the FISH protocol, optimized each of these, generated all of the figure panels not previously mentioned, and wrote the paper.

This work was published on December 26, 2012 in Proceedings of the National Academy of the United States of America (Beliveau et al., 2012)
Chapter 2 Acknowledgements

We would like to thank: S. Nguyen, C. Kim-Kiselak, and M. Hannan of the Wu lab, G. Church, F. Bantignies, J. Bateman, J. Birchler, R. Daniels, P. Ferree, J.T. Lee, A. Lindgren, C. Morton, J. Mosberg, Y. Murgha, J. Seidman, M. Sismour, S. Sun, S. Vassalo, Fre. Vigneault, Fra. Vigneault, H. Wakimoto, Y. Wakimoto, and L. Yang for advice and technical assistance; R. Jungman, T. Schmidt, W. Shih, and P. Yin for stimulating discussion; S. Clewely, C. Botka, and the RITG for computational assistance; M. Muscato and B. Schneider of Olympus and K. Stevens of IDT for generosity and technical expertise; S. Elledge, R. Kingston, D. Moazed, and M. Thomas for helpful feedback. This work was supported by an NIH/NIGMS grant (1RO1GM085169) and Pioneer Award, a Broad Institute SPARC Award, and a Cox Program Award from Harvard Medical School to C.-t.W., an NSF Graduate Fellowship to B.R.W, an NIH/NCI Ruth L. Kirschstein NRSA to E.F.J.. an NIH CEGS grant to G. Church in support of J.B.L., NIH grants (5R42GM097003, 1R43GM093579) to J.M.R. for the development of the oligo synthesis technology, and grant 1S10RR028832-01 to C. Botka and the RITG.
### Abbreviations used in Chapter 2

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2N</td>
<td>Diploid</td>
</tr>
<tr>
<td>2R</td>
<td>The right arm of Drosophila chromosome 2</td>
</tr>
<tr>
<td>4N</td>
<td>Tetraploid</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence \textit{in situ} hybridization</td>
</tr>
<tr>
<td>GC%</td>
<td>% guanine + cytosine</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked nucleic acid</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>NE</td>
<td>Nicking endonuclease</td>
</tr>
<tr>
<td>Oligo</td>
<td>Oligonucleotide</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PNA</td>
<td>Peptide nucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SSC</td>
<td>Saline sodium citrate buffer</td>
</tr>
<tr>
<td>SSCT</td>
<td>Saline sodium citrate buffer with Tween-20</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>$T_M$</td>
<td>Melting temperature</td>
</tr>
</tbody>
</table>
Chapter 2 Abstract

A host of observations demonstrating the relationship between nuclear architecture and processes such as gene expression have led to a number of new technologies for interrogating chromosome positioning. While some of these technologies reconstruct intermolecular interactions, others have enhanced our ability to visualize chromosomes in situ. Here, we describe an oligonucleotide- and PCR-based strategy for fluorescence in situ hybridization (FISH) and a bioinformatic platform that enables this technology to be extended to any organism whose genome has been sequenced. The oligonucleotide probes are renewable, highly efficient, and able to robustly label chromosomes in cell culture, fixed tissues, and metaphase spreads. Our method gives researchers precise control over the sequences they target and allows for single and multicolor imaging of regions ranging from tens of kilobases to megabases with the same basic protocol. We anticipate this technology will lead to an enhanced ability to visualize interphase and metaphase chromosomes.
Chapter 2 Introduction

The role of chromosome positioning in gene regulation and chromosome stability is fueling a growing interest in technologies that reveal the in situ organization of the genome. Among these technologies are chromosome conformation capture (3C) (Dekker et al., 2002) and its several iterations, such as Hi-C (Lieberman-Aiden et al., 2009), which are applied to populations of nuclei in order to identify chromosomal regions that are in close proximity to each other (de Wit and de Laat, 2012; Tanizawa and Noma, 2012). Another technology is fluorescence in situ hybridization (FISH), wherein nucleic acids are targeted by fluorescently labeled probes and then visualized via microscopy; this technology is an extension of methods that once used radioactive probes and autoradiography but have since been adapted to use nonradioactive labels (Pardue and Gall, 1969; Bauman et al., 1980; Levsky and Singer, 2003; Gilbert et al., 2005; Volpi and Bridger, 2008; Cremer T and Cremer M, 2010; Itzkovitz and van Oudenaarden, 2011). FISH is a single-cell assay, making it especially powerful for the detection of rare events that might be otherwise lost in mixed or asynchronous populations of cells. In addition, because FISH is applied to fixed cells, it can reveal the positioning of chromosomes relative to nuclear, cytoplasmic, and even tissue structures. FISH can also be used to visualize RNA, permitting the simultaneous assessment of gene expression, chromosome position, and protein localization.

FISH probes are typically derived from cloned genomic regions or flow-sorted chromosomes, which are labeled directly via nick translation or PCR in the presence of fluorophore-conjugated nucleotides or labeled indirectly with nucleotide-conjugated haptens, such as biotin and digoxigenin, and then visualized with secondary detection reagents. Probe DNA is often fragmented into ~150 - 250 bp pieces to facilitate its penetration into fixed cells (Lichter et al., 1988) and, as many genomic clones contain repetitive sequences that occur abundantly in the
genome, hybridization is typically performed in the presence of unlabeled repetitive DNA (Landegent et al., 1987). Another limitation to clone-based probes is that the genomic regions that can be visualized with them are restricted by the availability of clones and the size of their genomic inserts, which typically ranges from 50 - 300 kb. While it is possible to target larger regions and establish banding patterns by combining probes (Volpi and Bridger, 2008; Schröck et al., 1996; Speicher et al., 1996; Tanke et al., 1996; Shopland et al., 2006), this approach is often challenging, as each clone needs to be prepared and optimized for hybridization separately. The efficiency of these probes can also be variable, even among different preparations of the same probe. This variation may sometimes be a consequence of random labeling and fragmentation during probe production.

Many types of custom-synthesized oligonucleotides (oligos) have also been used as FISH probes, including peptide nucleic acid (PNA) and locked nucleic acid (LNA) oligos (Larsson et al., 1988; Lansdorp et al., 1996; Silahtaroglu et al., 2003; Pellestor et al., 2005; Müller et al., 2010; Briones and Moreno, 2012). Rather than relying on the isolation of a clone, such probes are designed to target precisely defined sequences. Also, as these probes are typically short (~20 - 50 bases) (Dernburg et al., 1996; O’Keefe et al., 1996; Femino et al., 1998) and single-stranded, they diffuse efficiently into fixed cells and tissues and are unhindered by competitive hybridization with complementary probe fragments. Oligo probes have allowed the visualization of single-copy viral DNA as well as individual mRNA molecules using branched DNA signal amplification (Player et al., 2001) or a handful to a few dozen short oligo probes (Femino et al., 1998; Raj et al., 2008), and, by targeting blocks of repetitive sequences as a strategy to amplify signal, enabled the first FISH-based genome-wide RNAi screen (Joyce et al., 2012). Oligo probes have also been generated directly from genomic DNA using parallel PCR reactions.
(Navin et al., 2006; Lamb et al., 2007). However, the high cost of synthesizing oligo probes has limited their use.

The availability of complex oligo libraries produced by massively parallel synthesis has enabled a new generation of oligo-based technologies. These libraries are synthesized on a solid substrate, then amplified or chemically cleaved in order to move the library into solution (Porreca et al., 2007; Gnirke et al., 2009). Two very recent studies have used complex libraries to visualize single-copy regions of mammalian genomes by FISH. One study used long oligos (>150 bases) as templates for PCR, and then labeled the amplification products non-specifically (Yamada et al., 2011), while the other adapted a 75-100 base single-stranded sequence-capture library for FISH by replacing the 5’ biotin with a fluorophore (Boyle et al., 2011).

Here we report a method, called 'Oligopaints', that utilizes oligo libraries as a renewable source of FISH probes carrying only 32 bases of homology to the genome. We amplify these libraries with fluorophore-conjugated PCR primers, thereby ensuring one fluorophore per oligo probe, and are, furthermore, able to process the amplification products enzymatically to produce highly efficient single-stranded, strand-specific probes that can visualize regions ranging from tens of kilobases to megabases. We also describe a set of bioinformatic tools to facilitate the design of these probes, which makes our technology compatible with any sequenced organism.

From oligo library to FISH probe

Our strategy for generating Oligopaints begins with the design and synthesis of libraries of single-stranded 74mers (ss74mers), where each oligo contains 32 bases of genomic sequence flanked by 21 base primer sequences (Figure 2.1). As such, more than one probe set can be
Figure 2.1 In this scheme, each oligonucleotide in the library is composed of 32 bases of genomic sequence flanked by 21 base primer sequences. One of the primers carries a 5’ fluorophore, while the other contains a recognition site for a nicking endonuclease (NE) (Xu et al., 2007). A nicking reaction followed by denaturing gel electrophoresis yields 53 base ssDNAs.

The oligos are then synthesized from the same library through the use of multiple primer pairs. The oligos are then amplified via PCR, which can be carried out with or without an emulsion (see Chapter 2 Materials and Methods) (Williams et al., 2006). Importantly, one of the primers contains a 5’ conjugated fluorophore, while the other contains the recognition site for a nicking endonuclease (NE) (Xu, 2007), which provides a strategy for making Oligopaints single-stranded. As shown in Figure 2.1, the NE recognition site is oriented such that the nick occurs immediately 3’ of the 32 bases of genomic sequence on the labeled strand. Upon denaturation, the nicked strand separates into 53- and 21-base fragments, while the undigested strand remains at 74 bases. Finally, we use denaturing gel electrophoresis to isolate and extract the labeled ss53mers, which can then be used as a strand-specific FISH probe.
Genome-scale probe design

The Oligopaints approach calls for identifying genomically unique sequences with desirable hybridization properties. To this end, we have created a bioinformatics pipeline that utilizes the program OligoArray, which simulates the thermodynamics of probe-target hybridizations and allows the user to specify several parameters, including melting temperature ($T_m$), percent G+C content (GC%), and sequences to avoid (Roulliard et al., 2003). Candidate probes are then assessed using the UNAfold package (Markham and Zuker, 2008) for the propensity to form secondary structures and verified using BLAST to have only a single genomic target (Altschul et al., 1997).

In order to use OligoArray to design FISH probes, we first assembled a sequence database of tiled 1 kb segments for each genome analyzed. While Oligopaints can theoretically be made to carry any length of homology to the genome, we elected to search for 32mer sequences, as this length is compatible with short array formats and gave us the densest coverage in our pilot searches for probe sequences (Supplementary Table A1.1). We searched for unique 32mers with a $T_m$ between 75 - 90°C, in order to select probe sequences whose hybridization with their targets would withstand stringent FISH wash conditions, and with a GC% between 35 - 80%, in order to increase the likelihood of amplification. A minimum spacing of 10 bases between probe sequences was imposed to minimize steric interference by adjacent probes during hybridization, and homopolymeric stretches of 5 or more A’s or T’s, or 4 or more G’s or C’s were avoided in order to maximize PCR fidelity and minimize spurious probe-probe interactions. The thermodynamics were simulated at 70°C, as this temperature mimics the most stringent conditions under which we anticipated performing FISH washes.
Table 2.1 The occurrence of probe sequences in the genomes of five eukaryotic organisms. For each, we present the genome assembly version, the haploid genome size in megabase pairs, the number of probes found in millions, the mean density of probe sequences per kilobase, and our estimated % coverage. The estimation of % coverage is the percentage of 250 kilobase windows in a given genome in which at least 500 probe sequences occur. Genome size corresponds to size of the genome assembly used (see Chapter 2 Materials and Methods).

<table>
<thead>
<tr>
<th>Organism</th>
<th>Assembly</th>
<th>Size Mbp</th>
<th># Probes x 10^6</th>
<th>Probes/ kb</th>
<th>Coverage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. elegans</td>
<td>ce6</td>
<td>100</td>
<td>1.10</td>
<td>10.9</td>
<td>100</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>dm3</td>
<td>140</td>
<td>1.79</td>
<td>12.8</td>
<td>95.4</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>tair10</td>
<td>119</td>
<td>1.45</td>
<td>12.2</td>
<td>98.1</td>
</tr>
<tr>
<td>M. musculus</td>
<td>mm9</td>
<td>2,655</td>
<td>30.1</td>
<td>11.4</td>
<td>94.1</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>hg19</td>
<td>3,096</td>
<td>29.9</td>
<td>9.7</td>
<td>90.3</td>
</tr>
</tbody>
</table>

Genomic target sites are abundant

We have used OligoArray to mine the C. elegans, D. melanogaster, A. thaliana, M. musculus, and human genomes for probe sequences and, despite differences in genome size and complexity, found an average density of ~10 per kb in each (Table 2.1). Furthermore, although there are many 1 kb segments in which we found no appropriate sequences, the overall distribution of probe sequences tends to be fairly uniform (Supplementary Fig. A1.1-A1.3). These observations suggest that most genomic regions will be amenable to Oligopaints. Indeed, our coverage ranges from 90.3% for the human genome to 100% for the C. elegans genome (Table 2.1). Oligopaints is supported by a website (http://genetics.med.harvard.edu/oligopaints) that hosts files detailing the genomic locations of all the target sites we have discovered as well as a suite of scripts and documentation to assist with the design of probe sets (Supplementary Figure A1.4). The website also provides tools and instructions that will allow researchers to use
OligoArray to search for targets using different parameters and to mine the genomes of additional organisms for probe sequences.

**Oligopaints robustly label interphase and metaphase chromosomes**

We have used Oligopaints to visualize single-copy DNA in a variety of cell lines, including Drosophila Kc167 (XXXX 4N) and human WI-38 (XX 2N) and MRC-5 (XY 2N) cells, and have found it to be applicable for a range of target sizes ([Supplementary Table A1.2, A1.3](#)). At the lower end, we have found that a 10 kb interval on human chromosome 4 (4p16.1) with very dense coverage (>18 probe sequences/kb) could be visualized with 200 oligos, producing at least one FISH focus in 90% of WI-38 nuclei, with 88% displaying two foci (n = 137; **Figure 2.2A**). This same interval could also be visualized in human metaphase spreads ([Supplementary Figure A1.5](#)). We obtained similar success when we extended coverage in this region to 52 kb using an additional 650 probes, with 98% of nuclei displaying at least one FISH focus and 86% displaying two foci (n = 126; **Figure 2.2B**). We have also visualized larger regions, such as a 2.1 Mb interval on the human X chromosome. Here, we used 20,020 probes and observed 100% of nuclei with one focus and 97% with two in WI-38 (XX) cells (n = 119) and 98% of nuclei with at least one focus in MRC-5 (XY) cells (n = 124; **Figure 2.2C,D**). Comparable efficiency (95%; n = 136) was observed for a probe set composed of 25,000 oligos targeting a 2.7 Mb region (50D1-53C7) of the right arm of the second chromosome of Drosophila (2R) (**Figure 2.2E**).

Because our bioinformatics platform allows us to specify custom hybridization patterns, such as multicolor banding, the effectiveness of Oligopaints can be extended to chromosomal regions on the order of tens of megabases or more. For example, we have used three color FISH to
Figure 2.2 Oligopaints efficiently label interphase human and Drosophila nuclei. (A) A 200 oligo probe set targeting 10 kb at human 4p16.1 was hybridized to WI-38 (2N) cells. (B) A probe set targeting the region shown in (A), but extended to 850 oligos targeting 52 kb was hybridized to WI-38 cells. (C, D) A 20,020 oligo probe set targeting 2.1 Mb at human Xq13.1 was hybridized to WI-38 cells (XX) (C) or MRC-5 (XY 2N) cells (D). (E) A 25,000 oligo probe set targeting 2.7 Mb at 50D1-53C7 on Drosophila 2R was hybridized to Kc167 (4N) cells. An enlarged image of the inset is shown beneath each micrograph. Below: labeling efficiencies presented as the percentage of cells that displayed at least one FISH focus (Supplementary Table A1.2). All probe sets were labeled with TYE563 (Cy3 mimic; red); DNA was identified with DAPI (blue). Scale bars: 5 µm. Images are maximum Z projections. Each micrograph was acquired using parameters optimized for entire fields of cells; thus, the sizes of the foci do not necessarily correlate with the sizes of the targeted regions.

visualize 7.6 Mb on the human X (Xq13.1-q21.1; Supplementary Table A1.3) in both WI-38 interphase nuclei (Figure 2.3A) and human primary metaphase spreads (Abbott Molecular) (Figure 2.3B) using 60,060 oligos, and 19.5 Mb of Drosophila 2R (41E3-60D14; Supplementary Table A1.3) using 180,000 oligos (Figure 2.3C). The same 180,000 oligo pool has been used to reveal chromosome packaging in a polytenized salivary gland nucleus (Figure 2.3D), while a 75,000 oligo subset of that pool has been combined with fluorescently labeled wheat germ agglutinin to provide an in situ rendering of the positioning of Drosophila chromosome 2R relative to the nuclear envelope (Wright, 1984) (Figure 2.3E).
Figure 2.3 Multicolor FISH with Oligopaints. (A, B) Three 20,020 oligo probe sets targeting adjacent regions at human Xq13.1, Xq13.2, and Xq13.3-q21.1 (Supplementary Figure A1.1A; Supplementary Table A1.3) were used to produce 3-color FISH images from WI-38 (XX) interphase (A) and primary metaphase (Abbott Molecular) (B) chromosomes. Probe sets were labeled with TYE563 (Cy3 mimic; red), TYE665 (Cy5 mimic; white), or 6-FAM (green), respectively. (C, D) A 180,000 oligo probe set was used to paint a multicolor banding pattern from 41E3 to 60D14 (Supplementary Figure A1.1B; Supplementary Table A1.3) on Drosophila 2R in interphase Kc167 (4N) nuclei (C) and salivary gland polytene chromosomes (D) with the following pattern: 41E3-44C4 (white (C) or blue (D); 25,000 TYE665 labeled oligos), 44C4-50C9 (green; 52,500 6-FAM labeled oligos), 50D1-53C7 (red; 25,000 TYE563 labeled oligos), 53C9-58B6 (green, 52,500 6-FAM labeled oligos), and 58D2-60D14 (white (C) or blue (D); 25,000 TYE665 labeled oligos). (E) Two probe sets were combined to span Drosophila 2R. One probe set was composed of 25,000 TYE563 (red) labeled oligos targeting 50D1-53C7, while the second was composed of 25,000 TYE665 (white) labeled oligos targeting 41E3-44C4 and 25,000 TYE665 (white) labeled oligos targeting 58D2-60D14. The nuclear envelope was stained with wheat germ agglutinin conjugated to Alexa Fluor 488 (green). (A), (C), and (E) are maximum Z projections, while (B) and (D) are single Z slices. DNA was identified with DAPI (blue for A–C and E, gray for D). Scale bars: 10 µm.
We have also found Oligopaints to be quite robust. They are compatible with a range of hybridization and wash conditions (Supplementary Figure A1.6-A1.8), work in the 40-800 nanomolar range (Supplementary Table A1.2), and are amenable to repeated rounds of...
hybridization (Figure 2.4A; Supplementary Figure A1.9), indicating that they will enable researchers to ‘walk’ along the lengths of chromosomes, especially if each step, or hybridization, were itself to involve multi-color FISH targeting several contiguous regions. In addition, our bioinformatics platform yielded probe sets that permitted a single hybridization step to visualize a 2.5 Mb region centered on the X-inactivation center (XIC) and the Xist RNA produced by this region (Brown et al., 1991) (Figure 2.4B; Supplementary Figure A1.10); while simultaneous visualization of both genomic regions and transcripts has been achieved previously (Chaumeil et al., 2008), it typically is a laborious process requiring sequential hybridizations in order to avoid cross-talk between the probe sets targeting DNA and those targeting RNA (Lee et al., 1996). We have also found Oligopaints to be suitable for conducting high-throughput FISH in 384-well plates (Joyce et al., 2012) (Figure 2.4C; Supplementary Figure A1.11), opening up the possibility of Oligopaints-based whole-genome RNAi and small molecule screens. Finally, we anticipate that Oligopaints may be useful for discerning chromatin structure; our probes covering megabase-sized regions have occasionally produced foci that display sub-structures, with some appearing spherical or “ball-like” and others more linear or “thread-like” (Figure 2.4D).

**Chromosome painting in whole-mounted Drosophila ovaries**

To assess the efficacy of Oligopaints *in vivo*, we turned to the Drosophila ovary and, using a probe set targeting the 2.7 Mb 50D1-53C7 region, demonstrated robust labeling of three cell types: oocytes containing pachytene chromosomes undergoing homolog pairing, polytene nurse cells, and somatic follicle cells (Figure 2.5A; Supplementary Table A1.2). For example, 100% of pachytene nuclei (n = 28) identified using an antibody against the nuclear synaptonemal complex protein C(3)G (Page and Hawley, 2001) were labeled, with 89% containing a single
focus and the remaining 11% displaying two closely positioned foci (≤1.0 µm apart), as would be expected for cells containing paired homologs. This result is especially notable, as condensed, pachytene chromosomes are often difficult to label. Our probes also labeled 100% of the polytene nurse cells (n = 24) and follicle cells (n = 110) (Supplementary Table A1.2). Moreover, we attained 99-100% efficient multi-color FISH in all three cell types by introducing a second probe set targeting the 3.1 and 2.6 Mb regions of 41E3-44C4 and 58D2-60D14, respectively, flanking 50D1-53C7 (Figure 2.5B-D; Supplementary Table A1.2). These results

Figure 2.5 Oligopaints efficiently label nuclei from whole-mounted Drosophila ovaries. (A) A cartoon of a Drosophila ovariole displaying three cell types: the pachytene oocytes and polytene nurse cells within the meiotic cysts, and the somatic follicle cells that encase them. (B) A Drosophila germarium labeled with two probe sets composed of 25,000 TYE563 (Cy3 mimic; red) labeled oligos targeting a 2.7 Mb region at 50D1-53C7 and an additional set composed of two pools of 25,000 TYE665 (Cy5 mimic; green) labeled oligos targeting 41E3-44C4 and 58D2-60D14 (green), all regions located on Drosophila 2R. An antibody to the synaptonemal complex component C(3)G (white) was used to identify oocytes. Hashed circles demarcate the meiotic cysts. (C, D) The same probe set as described for panel (B) in a magnified view of a single oocyte (C) and a polytene nurse cell (D). (E) The 50D1-53C7 probe set in a magnified view of follicle cells. Scale bars: 5 µm. For all panels, DNA was identified by DAPI (blue).
demonstrate that Oligopaints are capable of labeling chromosomes from whole-mounted tissue preparations with high efficiency regardless of the copy number or level of compaction.

**Chapter 2 Discussion**

Oligopaints are renewable and highly efficient probes that are amenable to studies of any sequenced organism. In addition to revealing chromosome positioning, the probes should facilitate the detection of chromosomal aberrations, especially in conjunction with combinatorial labeling technologies (Volpi and Bridger, 2008; Schröck et al., 1996; Speicher et al., 1996; Tanke et al., 1999), and therefore contribute also in the clinical setting. In terms of expense, discounting the one-time investment for an oligo library, the cost of consumables (**Supplementary Table A1.4**) ranges from ~$0.10 - $1.50 per assay and is significantly below the cost of commercial probes. Even so, we are continuing to work toward improving the yield and reproducibility of our probe preparations, and, hence, reducing cost. Importantly, probe sets targeting several megabases work at the same concentrations as do probe sets targeting tens of kilobases (**Supplementary Table A1.2**), making the cost of painting large stretches of the genome extremely low.

We are especially interested in the capacity of Oligopaints to reveal the telomere to telomere positioning of interphase chromosomes and believe that our short, single-stranded oligo probes are particularly suited for this task. For instance, clone-based FISH probe fragments are typically much longer and thus may bind to their genomic targets without being fully hybridized; this may place the label further from the chromosome as well as promote the formation of networks of interacting probe fragments that could extend well beyond their genomic target. In contrast, the short, strand-specific nature of Oligopaints argue that they are more likely to fully
hybridize to their target and less likely to network, ensuring that they 'hug' the chromosome and thereby enhance their ability to reveal chromosome structure. These features may also enhance entry of Oligopaints into the nucleus and their maneuverability through fixed chromatin. The advantages over conventional probes may be particularly relevant at the <100 nm resolution of super-resolution fluorescence microscopy (Rust et al., 2006; Huang et al., 2010; Jungmann et al., 2010; Flors and Earnshaw, 2011), wherein precision in probe placement and an exact number of fluorophores per probe could augment image interpretation.

Finally, we note that our technology is versatile and able to interface with many other technologies (Supplementary Figure A1.12). Importantly, the primer sequences retained in the probe provide general strategies for coupling a wide variety of functionalities to bioinformatically designed oligo libraries, thereby extending the potential usefulness of Oligopaints. Functionalities can be attached directly to primers and incorporated into the probe during amplification, or they can be brought in by the hybridization of a secondary oligo (Jungmann et al., 2010; Li et al., 2005; Lin et al., 2007) that is homologous to the primer sequence; either way, significant cost savings can be achieved by bulk orders of modified oligos that can then be applied to any number of libraries. The availability of primer sequences further opens opportunities for bringing in functionalities via DNA binding factors or assembling DNA structures, such as those used in branched signal amplification (Itzkovitz and van Oudenaarden, 2011; Player et al., 2001). Thus, we believe that Oligopaints have the potential to become a reagent not only for visualization, but also a broader spectrum of methods that require the targeting of biochemical modifications and functional chemistries to nucleic acids in a sequence-specific fashion.
Chapter 2 Materials and Methods

Genome sequences
The ce6, dm3, mm9, and hg19 genomic sequences were obtained from the UCSC genome bioinformatics website. The tair10 sequence was obtained the JGI & CIG Phytozome website.

Probe discovery
Genomes were inputted into OligoArray2.1 (Roulliard et al., 2003), which was run on the Harvard Medical School RITG Orchestra UNIX cluster with the following parameters: -n 22 -l 32 -L 32 -D 1000 -t 75 -T 90 -s 70 -x 70 -p 35 -P 80 -m "GGGG;CCCC;TTTTT;AAAAA" -g 42.

PCR primers
Please see Appendix 1, Supplementary Methods.

Oligonucleotide libraries
Oligonucleotide libraries were synthesized by MYcroarray (Ann Arbor, MI, USA). Libraries were either ordered as ssDNA 74mers or ssDNA 60mers. 60mer libraries were extended to 74mers using the emulsion PCR protocol detailed in (Williams et al. 2006) and using the “touch-up” cycle described for the generation of probe set for the Xist RNA.

Emulsion PCR amplification
Our strategies for PCR have evolved with the development of the technology, involving changes in template and primer concentrations and ratios as well the use of emulsion. Although we conducted all earlier PCR reactions for the preparation of probe with emulsion (Figures 2.2A-E; 2.3A,C-D; 2.4A,C-D; 2.5B-E; A1.6A,B; A1.7; A1.8; A1.9; A1.11) we now use this protocol
primarily for renewing the library (as versus for the generation of FISH probe). For more information about the protocol, please see Appendix 1, Supplementary Methods.

**Non-emulsion PCR amplification**

DNA FISH probe sets amplified without emulsion used the following cycle: 95°C for 5’, 3 cycles of 95°C for 30s, 50°C for 30s, 72°C for 15s, followed by 40 cycles of 95°C for 30s, 60°C for 30s, and 72°C for 15s, with a final extension step at 72°C for 5’. 100 pmoles of each primer and 1 ng of template were used per 100 µl of PCR reaction.

**FISH probe set targeting the Xist RNA**

The probe set for the Xist RNA was made from 96 ssDNA 60mers (IDT), each containing 32 bases of exonic sequence from Xist exon 1 flanked by 14 base primer sequences. The 60mers were then amplified in a pool using 21 base primers and the same cycle and parameters listed above for non-emulsion PCR amplification except that 10 femtomoles of template were used per 100 µl of PCR reaction.

**Extraction and purification of ssDNA**

Please see Appendix 1, Supplementary Methods.

**Slide preparation for interphase FISH**

Glass slides (Thermo 4951-001) were treated with a 0.01% (v/v) poly-L-Lysine solution (Sigma P8920) for 5’, then air-dried. 100 µl of a 1-2 x 10^6 cells/ml solution was spotted on each slide and allowed to adhere for 1-2 hrs at 23°C (Drosophila) or 37°C (mammalian). Slides were rinsed in 1X PBS at room temperature (RT), then fixed at RT for 5’ or 15’ in 1X PBS + 4% (v/v) paraformaldehyde (Electron Microscopy Sciences 15710). Post fixation, slides were rinsed in 1X
PBS, washed for 5' in 2X SSCT (0.3 m NaCl, 0.03 m NaCitrate, 0.1% Tween-20), then washed for 5' in 2X SSCT + 50% formamide (v/v), all at RT. Slides were then transferred to fresh 2X SSCT + 50% formamide for storage at 4°C. For information on cell culture, see Appendix 1, Supplementary Methods.

**Interphase FISH**

Slides were warmed to RT, then incubated for 2.5' or 3' in 2X SSCT + 50% formamide (v/v) at 92°C, then incubated for 20' at 60°C in 2X SSCT + 50% formamide, then cooled to RT. 1-20 pmole of each probe was then added to each slide as part of a 25 μl hybridization cocktail composed of 2X SSCT, 50% formamide, 10% dextran sulfate (w/v), and 10 μg of RNase A (Fermentas EN0531) and sealed beneath a 22 x 22 cm #1.5 coverslip using rubber cement. Slides were denatured for 2.5' at 92°C on a water-immersed heat block, then allowed to hybridize overnight at 37°C or 42°C in a humidified chamber. Slides were washed for 15' in 2X SSCT at 60°C, then for 10' 2X SSCT at RT, and then for 10' in 0.2X SSC at RT. Slides were then mounted in SlowFade Gold + 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen S36938) and sealed with a 22 x 30 #1.5 cm coverslip using nail polish.

**Simultaneous RNA/DNA FISH**

Simultaneous RNA/DNA FISH was performed using a “3D-FISH” protocol adapted from (Lanzuolo et al., 2007). For more information, please see Appendix 1, Supplementary Methods.

**Metaphase spreads**

(46, XY) metaphase spreads were obtained from Abbott Molecular (30-806010).
Metaphase FISH protocol

Please see Appendix 1, Supplementary Methods.

Rehybridization protocol

The same FISH protocol was used as for single-round interphase FISH, with probe being stripped off between hybridizations by a 40 second wash in 2X SSCT + 50% formamide (v/v) at 65°C.

Modified 384-well FISH protocol for Oligopaints

Please see Appendix 1, Supplementary Methods.

Hybridization to Drosophila ovarioles and polytene chromosome squashes

Please see Appendix 1, Supplementary Methods.

Microscopy and image processing

Please see Appendix 1, Supplementary Methods.
Chapter 3:

In situ single-molecule localization super-resolution imaging and haplotype visualization using Oligopaint FISH probes
Chapter 3 Contributions

The work presented in this chapter would not be possible without wonderful collaborations with members of the C.-ting Wu lab as well as members of the laboratories of Xiaowei Zhuang, Peng Yin, William Shih, and Jeannie Lee. In particular, Alistair Boettiger of the Zhuang laboratory performed all of the STORM imaging and polymer modeling presented (Figure 3.2, Supplementary Figure A2.11) and wrote the corresponding methods section and some of the corresponding body text. Maier Avendaño and Ralf Jungmann of the Peng Yin laboratory performed all of the DNA-PAINT imaging presented (Figure 3.3, Supplementary Figure A2.12) and wrote the corresponding methods section and some of the corresponding body text (R.J. is joint supervised by P.Y. and W.S.). David Colognori of the Jeannie Lee lab shared his 129S1/SvImJ and CAST/EiJ fibroblast cell lines, which were instrumental to our development of the HOPs probes. Chamith Fonseka and Ruth McCole of the Wu laboratory developed the snpPopper.py script necessary to design HOPs probes. Ruth also performed the foundational work on the DGRP variant data that enabled the Drosophila HOPs to be designed and ordered. Caroline Kim and Frederic Bantignies of the Wu laboratory helped to develop and optimize the secondary oligo system (work performed while F.B. was on sabbatical). Frederic Bantignies also pioneered the use of locked nucleic acids as secondary oligos. Eric Joyce of the Wu laboratory performed FISH on Drosophila tissues (Figure 3.4E, Supplementary Figure A2.17, A2.18) and embryos (Figure 3.4F), with the latter work being performed in conjunction with Jelena Erceg of the Wu Laboratory. Eric also developed the lambda exonuclease probe generation method (Supplementary Figure A2.3), and wrote the corresponding methods section. Mohammed Hannan and Hien Hoang of the Wu laboratory assisted with the synthesis of Oligopaint probe sets and cell culture. C.ting Wu supervised the project, provided funding, helped to design and analyze experiments, co-wrote the manuscript, and provided invaluable
support and mentorship. I designed, performed, and analyzed all proof-of-principle experiments for the secondary oligo system, conceived of using the secondary oligo system for dye pairing in STORM, developed the 3D-FISH protocols for STORM and DNA-PAINT, performed all of the FISH for STORM and DNA-PAINT, created the HOPs system, designed, performed, and analyzed all HOPs experiments in mouse cell culture, performed the experiments that produced every figure panel and table not previously mentioned, and wrote the manuscript.

This work is currently in review awaiting our resubmission at *Nature Communications*. 
Chapter 3 Acknowledgements

We would like to thank members of the Wu lab, K. Ahmad, G. Orsi, T. Schmidt, S. Elledge, R. Kingston, D. Moazed, D. Zhang, R. Oakey, M. Cowley, N. Apostolopoulos, and L. Cai for helpful discussions; B. Harada, P. Huang, G. Church, F. Winston, S. McCarroll, D. Reich, J. Seidman, and C. Seidman for discussion, equipment, and technical assistance; S. Clewley and J.H. McDonald for computational assistance. Finally, we offer belated appreciation to Rachel O’Neill and Judy Brown, who tested our original Oligopaints technology when it was first developed.

BJB, RBM, EFJ, CK-K, FB, CYF, JE, MAH, HGH, and CtW were supported by awards from NIH/NIGMS (RO1GM61936, 5DP1GM106412) and Harvard Medical School (HMS) to CtW, NIH/NCI (F32CA157188) to EFJ, HMS to CK-K, and the Centre National de la Recherche Scientifique and the Fulbright Visiting Scholar Program to FB. ANB and XZ were supported by awards from the NIH and HHMI to XZ and the Damon Runyon Cancer Research Foundation to ANB. RJ, MSA, and PY were supported by awards from NIH (1DP2OD007292, 1R01EB018659, 5R21HD072481), ONR (N000141110914, N000141010827, N000141310593), and NSF (CCF1054898, CCF1162459) to PY. RJ and WMS were supported by an award from from NIH (1DP2OD004641) to WMS. PY and WMS were also supported by the Wyss Institute for Biologically Engineering. RJ and MSA were also supported by the Alexander von Humboldt-Foundation and the HHMI, respectively. DC and JTL were supported by an NIH grant, RO1-GM090278, and the HHMI.
### Abbreviations used in Chapter 3

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</td>
<td>129S1/SvImJ</td>
</tr>
<tr>
<td>BX-C</td>
<td>Bithorax complex</td>
</tr>
<tr>
<td>CAST</td>
<td>CAST/EiJ</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-coupled device</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNA-PAINT</td>
<td>DNA-based point accumulation in nanoscale topography</td>
</tr>
<tr>
<td>EMCCD</td>
<td>Electron-multiplying charge-coupled device</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>HILO</td>
<td>Highly inclined and laminated optical sheet</td>
</tr>
<tr>
<td>HOP</td>
<td>Homolog-specific oligopaints</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked nucleic acid</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblast</td>
</tr>
<tr>
<td>Oligo</td>
<td>Oligonucleotide</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PBT</td>
<td>Phosphate buffered saline with Tween-20</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SIM</td>
<td>Structured illumination microscopy</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSC</td>
<td>Saline sodium citrate buffer</td>
</tr>
<tr>
<td>SSCT</td>
<td>Saline sodium citrate buffer with Tween-20</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>STORM</td>
<td>Stochastic optical reconstruction microscopy</td>
</tr>
<tr>
<td>TIRF</td>
<td>Total internal reflection fluorescence</td>
</tr>
<tr>
<td>XIC</td>
<td>X-inactivation center</td>
</tr>
</tbody>
</table>
Chapter 3 Abstract

Fluorescence in situ hybridization (FISH) is a powerful single-cell technique for studying nuclear structure and organization. Here, we report two advances in FISH-based imaging. We first describe the in situ visualization of single-copy regions of the genome using two single-fluorophore super-resolution methodologies. We then introduce a robust and reliable system that harnesses single nucleotide polymorphisms (SNPs) to visually distinguish the maternal and paternal homologous chromosomes in mammalian and insect systems. Both of these new technologies are enabled by renewable, bioinformatically-designed, oligonucleotide-based “Oligopaint” probes, which we augment here with a strategy that uses secondary oligonucleotides (oligos) as a means to produce and enhance fluorescent signals. Our advances will substantially expand the ability of researchers to query parent-of-origin specific chromosome positioning and gene expression on a cell-by-cell basis.
Chapter 3 Introduction

Since their inception (Pardue and Gall, 1969; van der Ploeg, 2000; Levsky and Singer, 2003), in situ hybridization techniques have provided critical insights into the spatial organization of nucleic acids within the cell. This family of methodologies has led to the discovery that the eukaryotic nucleus is a highly ordered compartment, with chromosomes falling into distinct territories (Bolzer et al., 2005). Yet, despite decades of advances in hybridization-based single-cell imaging technology, our ability to directly visualize the fine-scale structure of the in situ genome remains constrained by the optical resolution of light microscopy and the limitations of our ability to target regions of interest. Consequently, many gaps remain in our understanding of how local chromatin structure and nuclear positioning impact processes such as transcription, the establishment of chromosome-chromosome interactions, and DNA repair.

Here, we report two strategies for in situ single-cell imaging, one that facilitates single-molecule super-resolution and another that utilizes SNPs to visually distinguish homologous chromosomes. Both take advantage of Oligopaints, which are highly efficient, renewable, strand-specific fluorescence in situ hybridization (FISH) probes derived from complex single-stranded DNA (ssDNA) libraries (Figure 3.1A). In contrast to classical FISH probes, which are produced from segments of purified genomic DNA amplified in bacterial vectors or PCR reactions, Oligopaints belong to a new generation of probes that are derived entirely from synthetic DNA oligonucleotides (oligos) (Yamada et al., 2011; Boyle et al., 2011; Beliveau et al., 2012). Such probes have the advantage of having their sequences chosen bioinformatically; thus, they can be designed to target any organism whose genome has been sequenced, engineered to avoid repetitive elements, and selected to have specific hybridization properties.
Central to the design of Oligopaints is the inclusion of non-genomic sequences flanking the region of homology to the genome, as these sequences enable amplification by PCR or other methods to produce DNA or RNA oligos, introduction of label, and conversion of double-stranded to single-stranded products (Beliveau et al., 2012) (Figure 3.1A). This design also permits the multiplexing of Oligopaint libraries, wherein a single library is used to produce multiple distinct probe sets, each derived from a subset of oligos through amplification via a primer pair specific for that subset (Beliveau et al., 2012) (Supplementary Figure A2.1). Furthermore, as a non-genomic sequence is designed to remain single-stranded when Oligopaint probes are hybridized to their genomic targets, it could be used to recruit activities without disruption of targeting. Indeed, the non-genomic sequence, which we call MainStreet (Beliveau et al., 2012), could be populated by any number of functionalities via the binding of complementary oligos, nucleic acid binding proteins, or other factors.

Results

We began our current studies by examining the ability of MainStreet to recruit a fluorescently labeled ‘secondary’ oligo (Figure 3.1B), as we were intrigued by the potential of secondary oligos to simplify the use of multiplexed Oligopaint libraries. In particular, if a common secondary oligo binding site were introduced into all of the probe sets of a library, all the probe sets could then be indirectly labeled in situ with the same species of secondary oligo. Such a strategy would remove the need to incorporate fluorophores directly into the Oligopaint probes, thereby reducing the number and, hence, cost of fluorophore-labeled oligos needed to utilize heavily multiplexed libraries (Supplementary Figure A2.1).
Figure 3.1

Secondary oligos are specific and efficient. (A) Synthesis strategy for Oligopaints, in which complex ssDNA libraries are amplified, labeled, and then purified as ssDNA probes (adapted from [Beliveau et al., 2012]; also see Supplementary Figure A2.2,A2.3). The amplification step is enabled by primer sequences contained within MainStreet, which consists of nongenomic sequences that will form ssDNA overhangs when Oligopaint probes are hybridized to their target. (B) A secondary (2°) oligo probe hybridized to MainStreet. Here, the secondary oligo carries a single, 5’ fluorophore that matches the fluorophore present on the Oligopaint (primary) probe, but in practice the number, identity, and placement of fluorophores on the secondary oligo can vary. (C) Two-color co-localization experiments in diploid Drosophila clone8 cells and diploid human WI-38 cells. The genomic target, number of nuclei examined (n), percent of nuclei (% Labeling) that had at least one signal from the primary (1°, Oligopaint) probe and at least one signal from the secondary oligo, and percent primary signals that have an overlapping secondary signal (%Co-localization) are given for each experiment; target size is shown in panel D. (D) Two-color co-localization, target location, and target size are shown for each secondary oligo in Drosophila clone8 (top) or human WI-38 (bottom) cells. DNA is stained with DAPI (blue). Images are maximum Z projections from a wide-field epifluorescent microscope. Note that our widefield microscope has an uncorrected chromatic aberration that results in the red channel being displayed ~170 nm to the right of the green channel (B. Beliveau and C.-ting Wu, unpublished data), which explains the shift observed in (D).
Figure 3.1 (Continued)

A

Complex ssDNA library → Amplification, ssDNA isolation → Oligopaint probes → Genomic homology → MainStreet (nongenomic) → FISH

B

32 base 2° oligo

Binding site for 2° oligo

Region of genomic homology

C

<table>
<thead>
<tr>
<th>2° oligo</th>
<th>Target</th>
<th>n</th>
<th>% Labeling</th>
<th>% Co-localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 89D-89E</td>
<td>139</td>
<td>99, 99</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2 89B-89D</td>
<td>113</td>
<td>100, 100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>3 89B-90D</td>
<td>142</td>
<td>98, 98</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>4 89B-89D</td>
<td>126</td>
<td>98, 98</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>5 89D-89E</td>
<td>115</td>
<td>100, 100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>6 89D-89E</td>
<td>126</td>
<td>100, 100</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th>n</th>
<th>% Labeling</th>
<th>% Co-localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>19q13.11-q13.12</td>
<td>111</td>
<td>100, 100</td>
<td>100</td>
</tr>
<tr>
<td>19q13.11-q13.12</td>
<td>121</td>
<td>100, 100</td>
<td>100</td>
</tr>
<tr>
<td>19q13.32-q13.33</td>
<td>126</td>
<td>100, 100</td>
<td>100</td>
</tr>
<tr>
<td>19q13.32-q13.33</td>
<td>112</td>
<td>100, 100</td>
<td>100</td>
</tr>
<tr>
<td>19q13.2-q13.31</td>
<td>120</td>
<td>100, 100</td>
<td>100</td>
</tr>
<tr>
<td>19q13.2-q13.31</td>
<td>144</td>
<td>100, 100</td>
<td>100</td>
</tr>
</tbody>
</table>

D

Drosophila clone 8

| 2° oligo 1 | 316 kb | 10 μm |
| 2° oligo 2 | 176 kb | 10 μm |
| 2° oligo 3 | 176 kb | 10 μm |

Human WI-38

| 2° oligo 4 | 2.3 Mb | 10 μm |
| 2° oligo 5 | 2.1 Mb | 10 μm |
| 2° oligo 6 | 2.1 Mb | 10 μm |
Implementing secondary oligos

We first used a database of orthogonal sequences (Xu et al., 2009) to design six 32-base DNA oligos with thermodynamic properties similar to those of our Oligopaint probes (Supplementary Table A2.1). Then, using touch-up PCR (Ailenberg and Silverman, 2000), we placed a binding site for one or more of the secondary oligos 5’ of the primer sequences in MainStreet (Figure 3.1B); this strategy allows binding sites to be added to any existing Oligopaint library and is compatible with both our published probe synthesis protocols (Beliveau et al., 2012; Beliveau et al., 2014) (Supplementary Figure A2.2, Chapter 3 Materials and Methods) as well as alternative methods for generating Oligopaints, such as our one-day method using lambda exonuclease (Little, 1981) (Supplementary Figure A2.3) and the MYtags strategy (MYcroarray, Ann Arbor, MI). Binding sites for secondary oligos could also be incorporated during the original design of the library, in which case they would be internal to the primer pairs specific to each probe set.

To assess the effectiveness of secondary oligos, we conducted two-color co-localization experiments in Drosophila and human cell culture. In these experiments, Oligopaint probe sets targeting regions ranging in size from 176 kilobases to 3 megabases and consisting of thousands to tens of thousands of oligos, each bearing a 5’ fluorophore as well as a binding site for a secondary oligo, were co-hybridized with a secondary oligo carrying a spectrally distinct fluorophore. We found all six of our secondary oligos to be remarkably specific, with 100% of the signals coming from the secondary oligos co-localizing tightly with the signals of the primary Oligopaint probes in both Drosophila and human cells (n >100 for all cases; Figure 3.1C,D and Supplementary Figure A2.4). The two-color FISH was also efficient, with 98-100% of nuclei (n >100 for all cases) displaying signals (Figure 3.1C). The secondary oligos can be added
simultaneously (Figure 3.1C,D) or sequentially (Supplementary Figure A2.5) and produce only weak speckling when they are added in the absence of primary probes. We observed a similarly robust performance when using 14-base secondary oligos containing locked nucleic acid (LNA) (Silahtaroglu et al., 2003) residues (Supplementary Table A2.1). Here, we used a single synthetic oligo, carrying a 32-base MainStreet and targeting the highly repetitive 359 satellite on the Drosophila X chromosome (100% co-localization, >99% efficiency for each of 3 LNAs, n >100 in all cases, Supplementary Figure A2.6); these LNA secondary oligos can either be directly labeled or programmed to form branched structures that amplify signals (Player et al., 2001) (Supplementary Figure A2.7-A2.9). In sum, our results suggest that secondary oligos hybridize efficiently to MainStreet and do not hinder the ability of Oligopaint probes to associate with their genomic targets, suggesting that MainStreet could also be used to augment the number of fluorophores at a genomic target via the recruitment of multiple secondary oligos, enable the combinatorial use of different fluorophores, and support applications involving Förster resonance energy transfer (FRET) (Blanco et al., 2009) (Supplementary Figure A2.10).

Enabling super-resolution FISH with Oligopaints and secondary oligos

The efficacy of secondary oligos raised the potential of their application for super-resolution microscopy (Huang et al., 2010; Flors and Earnshaw, 2011). As diffraction limits the resolution of conventional light microscopy to a distance of ~200 nm in the X-Y plane and ~500 nm in the Z direction, the volume of a diffraction-limited signal is considerably larger than that of many nuclear structures. In contrast, super-resolution imaging can resolve distances on the order of tens of nanometers (Huang et al., 2010; Xu et al., 2012; Vaughn et al., 2012) (Figure 3.2A). One family of super-resolution technologies relies on stochastically occurring single-molecule
fluorescence events, which are used to localize the position of each fluorophore molecule with high precision. These single-molecule-based super-resolution techniques can enhance our understanding of nanoscale structural features, as their resolution is limited only by the number of photons collected per fluorophore and the density at which the target structure is labeled with fluorophores (Huang et al., 2010). Importantly, they increase resolution by an order of magnitude over that provided by structured illumination microscopy (SIM) (Gustafsson, 2000), which has been the most broadly used super-resolution method to date for imaging genomic loci in situ (Nora et al., 2012; van de Corput et al., 2012; Patel et al., 2013; Smeets et al., 2014; Giorgetti et al., 2014). Yet, despite demonstrations that single-molecule-based super-resolution microscopy can be combined with FISH to visualize highly repetitive segments of the mammalian genome, such as constitutive heterochromatin (Müller et al., 2010; Weiland et al., 2011) and telomeres (Doksani et al., 2013), this approach has yet to be combined with DNA FISH to visualize genomically unique, non-repetitive sequence regions in situ.

We hypothesized that Oligopaint FISH probes may be well suited for super-resolution imaging, as they allow direct control over the number, position, and placement of fluorophore molecules on each Oligopaint oligo as well as those on any secondary oligos hybridized to MainStreet. Furthermore, our ability to control the length, orientation, and positioning of secondary oligos along MainStreet allows for the reliable placement of the fluorescent signal directly at the site of hybridization (Supplementary Figure A2.10), making them an ideal tool for tracing genomic structure at high resolution. This in mind, we first set out to explore the potential of combining Oligopaint probes with Stochastic Optical Reconstruction Microscopy (STORM) (Rust et al., 2006), which relies on the stochastic activation and localization of individual photo-switchable fluorophores to produce super-resolution images (Bates et al., 2005).
Figure 3.2 Super-resolution imaging with Oligopaints and STORM. (A) Schematic illustrating how a diffraction-limited FISH signal presents as many smaller fluorescence localizations via STORM (B) Simulated STORM images of two polymer models (left) illustrating the importance of localization density in resolving structure (total localizations in upper right corners). The color code on the polymer models traces along the length of the polymer (black to red to white). (C) Average number of localizations (mean + S.E.M.) per BX-C locus in Drosophila clone8 cells when the unlabeled primary probe is paired with a secondary oligo carrying Cy5 (left), when both the primary probe and secondary oligo carry Cy5 (middle), and when the primary probe carrying an AlexaFluor 405 activator is paired with a secondary oligo carrying Cy5. (D) Conventional (left) and STORM (right) images of the BX-C locus from three cells, with cell shown in bottom row exhibiting two loop-like protrusions. Right two panels: zoomed-in views of the boxed regions. (E) Simulation in which two-thirds of the localizations shown in image (D) have been removed at random to illustrate the loss of connectivity and structure in regions represented by a low density of localizations.
In this study, we used the photo-switchable cyanine dye Cy5 for STORM imaging. Cy5 can exist in two states: a ‘bright’ state, where it emits fluorescence upon excitation, and a ‘dark’ state, where it is not capable of fluorescing. Importantly, activation of Cy5 from the dark to the bright state can be enhanced by a nearby ‘activator’ dye. For instance, use of AlexaFluor 405 as the activator dye allows for photo-switching to be induced with an intensity of 405 laser excitation that is lower than that which would be used in the absence of activator dye, which may in turn allow the switching rate to be high while the 405 induced photobleaching rate remains low. In such an instance, more localizations can be recorded, thus improving the sampling resolution of the image. To explore the potential effects of localization density on resolution for chromatin structures, we simulated STORM images from hypothetical polymer structures (Figure 3.2B). We found that simulations with a low number of total localizations appeared more frequently as disconnected objects; while densely coiled parts of the polymer appeared similar across a broad range of total localizations, long protrusions and narrow bridges became difficult to distinguish from low levels of background when the number of total localizations was small.

We next harnessed our ability to create precise fluorophore-fluorophore pairings with Oligopaints and secondary oligos (Supplementary Figure A2.10), targeting the developmentally regulated 316 kb Drosophila bithorax complex (BX-C) (Lewis, 1998; Lanzuolo et al., 2007; Mallo and Alonso, 2003) for visualization. In particular, we paired Cy5 labeled secondary oligos with a primary probe set that carried either no label, a Cy5, or an AlexaFluor 405. Excitingly, all three primary-secondary pairings were able to produce super-resolution FISH images (Supplementary Figure A2.11). While all three primary-secondary pairings were effective, we observed a significantly greater number of single-molecule localizations when an AlexaFluor 405 activator dye with was paired the Cy5 reporter (median ± s.e.m: 2,075 ± 49, n = 434 for unlabeled primary/Cy5 labeled secondary; 3,364 ± 114, n = 133 for Cy5/Cy5; 5,612
±167, n = 353 for A405/Cy5, P = 1 x 10-23 for A405 vs. Cy5 and P = 1 x 10-90 for A405 vs. no label, Wilcoxon rank-sum test; Figure 3.2C, Supplementary Figure A2.11), demonstrating the effectiveness of dye pairing enabled by secondary oligos. The less than double the number of localizations observed with two Cy5 dyes per probe as versus a single Cy5 dye per probe is likely the result of quenching interactions between the reporter dyes. By taking advantage of the higher density of localizations made possible through the activator-reporter labeling strategy, we were further able to detect fine-scale nanostructures of chromatin, such as the one shown in Figure 3.2D, which is not visible in the diffraction-limited image of the same field. Indeed, while we find the BX-C locus in most cells to lack substantial protrusions, we did occasionally observe threads of chromatin appearing to loop away from the primary cluster of signals. Importantly, we found that if we approximate the labeling density obtained with a single Cy5 dye by removing two-thirds of localizations from our images at random, the shapes of the protrusions are not as clear (Figure 3.2E), with some segments becoming more difficult to distinguish from background (Supplementary Figure A2.11).

We also explored the potential of Oligopaint primary-secondary pairings to enable the visualization of single-copy genomic regions using a related single-molecule-based super-resolution approach called DNA-based Point Accumulation for Imaging in Nanoscale Topography (DNA-PAINT) (Sharonov and Hochstrasser, 2006; Jungmann et al., 2010; Jungmann et al., 2014). In DNA-PAINT, the single-molecule fluorescence events are generated by the transient hybridization of fluorescently labeled oligonucleotides, called “imager strands”, present in solution in the imaging buffer to complementary strands, called “docking strands”, on the target to be imaged, reminiscent of the binding of secondary oligos to the MainStreet of Oligopaints (Figure 3.3A); as the duplexes that form are designed to be unstable at room temperature (duplex length: 9 bases; bound time in imaging conditions ≈ 600 ms [Jungmann et
al., 2010], the transient binding interactions lead to an apparent “blinking” of the docking sites when imaged using total internal reflection fluorescence (TIRF) or a highly inclined and laminated optical sheet (HILO) (Tokunaga et al., 2008) configuration (Figure 3.3B).

Figure 3.3 Super-resolution imaging with Oligopaints and DNA-PAINT. (A) Labeling scheme using Oligopaint probes carrying an AlexaFluor 488 dye and a 9-base docking site that is complementary to imager strands labeled with ATTO 655. (B) Graph showing how the amount of time an imager strand is bound to its docking site translates into blinking. (C) Diffraction-limited image of an Oligopaints probe set labeled with AlexaFluor 488 and targeted to the mouse hoxB locus (left, single Z slice) and zoomed-in images of boxed regions using conventional microscopy (D1 and E1) and DNA-PAINT super-resolution microscopy (D2 and E2). The increase in spatial resolution is clearly visible.
In order to explore the feasibility of enabling DNA-PAINT imaging of chromosomes with Oligopaints, we designed a probe set consisting of 1,691 oligos carrying a binding site for an imager strand and targeting the developmentally regulated 174 kb hoxB locus (Mallo and Alonso, 2003) in mouse. Application of this probe set to primary mouse embryonic fibroblasts (MEFs) (Figure 3.3C) produced super-resolution images, wherein single molecule fluorescence events could be localized with an average precision of 8.6 nm, resulting in an obtainable resolution of ~20 nm. Note that we were able to maintain a constant number of single-molecule localizations per frame over the entire course of image acquisition because, as imager strands are continuously replenished from solution, photobleaching does not present a significant problem for DNA-PAINT (Supplementary Figure A2.12). Together with the STORM results, these data demonstrate that Oligopaints are a powerful tool for visualizing single-copy genomic loci using super-resolution imaging.

Distinguishing homologous chromosomes with Oligopaints

While the methods described above can enhance our capacity to resolve chromosomal structures, they do not address one of the most intractable challenges in single-cell studies of chromosome positioning and gene expression, which is the visual distinction of maternal, paternal, and, indeed, any homologous chromosomes (homologs). Strategies for distinguishing homologous chromosomes and chromosomal regions would greatly advance our capacity to investigate phenomena such as X-inactivation (Jeon et al., 2012), imprinted gene expression (Bartolomei and Ferguson-Smith, 2011), and random mono-allelic expression (Chess, 2012); the few methods that are available either rely on relatively inefficient enzymatic signal amplification strategies (Zhong et al., 2001; Larsson et al., 2004; Grundberg et al., 2013) or are appropriate only for highly repetitive portions of the genome (Nilsson et al., 1997) or RNA
molecules (Ohno et al., 2001; Hansen and van Oudenaarden, 2013; Levesque et al., 2013), and thus cannot be used to visualize single-copy regions or loci that are not expressed in the sample of interest. We have addressed this challenge by developing Homolog-specific OligoPaints, or ‘HOPs.’

HOPs take advantage of the abundant and well-characterized single nucleotide polymorphism (SNP) data, such as those provided by the Wellcome Trust Sanger Mouse Genomes Project (Keane et al., 2011) and the Drosophila Genetic Reference Panel (DGRP) (Mackay et al., 2012). In our approach, we first generate short blocks of reference genomic sequence centered on each SNP in the region we wish the HOPs to target (Supplementary Figure A2.13). We then input these blocks into our Oligopaint probe discovery pipeline (Beliveau et al., 2012) to identify probe sequences that overlap the location of at least one SNP, are genomically unique, and have suitable thermodynamic properties. Finally, we run a custom Python script to insert the SNP variant(s) into the probe sequences. Importantly, HOP probe sets are always made in pairs; that is, each oligo of a HOP probe set has a cognate oligo in its partner probe set, where both oligos span precisely the same genomic coordinates and differ only by the SNP variant(s) they carry. Thus, partner HOP probe sets target the same region on different homologs by utilizing differences in the haplotypes of these chromosomes.

In our first test of the HOPs system, we examined a 2.6-megabase region containing the murine X-inactivation center (XIC), which produces the Xist RNA (Jeon et al., 2012), in three SV-40 large T-antigen immortalized MEF lines (Figure 3.4A). These lines, all of which appear to carry four copies of the X chromosome, are derived from three strains of mice: 129S1/SvlmJ (129), CAST/EiJ (CAST), and hybrid 129xCAST mice (Yildirim, et al., 2011). Importantly, the 129 and CAST genomes differ by an average of 2-3 SNPs per kilobase both in the 2.6 Mb region of the
Figure 3.4 Homolog-specific OligoPaints (HOPs). (A) Schematic of HOPs targeting the mouse X-inactivation center (XIC; not to scale). 129 (green) and CAST (magenta) HOPs are targeted to SNPs and carry variants specific for the 129S1/SvImJ (129) or CAST/EiJ (CAST) genomes, respectively, while interstitial (white) probes target sequences common to both genomes. None of these three probe sets target the Xist transcript, which is targeted by a fourth Oligopaint probe set (blue) (B) Hybrid EY.T4 129xCAST transformed MEF cells visualized with 129 (green) and Cast (magenta) HOPs and the interstitial probe set (white). The interstitial probe set binds 129 and CAST chromosomes equally well (left), while the 129 and CAST HOPs reveal the parent-of-origin of the interstitial signals (right). (C) RNA/DNA FISH with 129 (green) and CAST (magenta) HOPs and Xist RNA FISH (white) demonstrating co-localization of Xist signal with that of the 129 HOP. Arrows point to Xist signals. (D) Percentage of nuclei falling into each of five Xist staining patterns. (E) Polytene chromosomes of a Drosophila salivary gland nucleus (left) and enlarged image of boxed region (right) from DGRP 057 x DGRP 461 hybrid larvae visualized with Oligopaints targeting the BX-C (blue) and 057-specific (green) and 461-specific (magenta) HOPs targeting the flanking 89E-93C region. DNA is stained with DAPI (grey), which is removed from right image. Images are single Z slices from a laser scanning confocal microscope. (F) Drosophila 6-8 hour embryo nuclei visualized with the BX-C probe set (white) and the 057 (green) and 461 (magenta) HOPs showing pairing (left) and unpairing (right) at both BX-C and the adjacent 89E-93C region. (G) % pairing observed at BX-C and 89E-93C, where loci were considered paired if edge-to-edge distance between their signals was ≤0.8 μm. (n.s., not significant, two-tailed Fisher’s exact P = 0.88, n = 101). (H) The pairing status of BX-C is statistically associated with that of 89E-93C (two-tailed Fisher’s exact P = 6.4 x 10^-17, n = 101). For (B), (C), and (F): DNA is stained with DAPI (blue). Images are maximum Z projections from a laser scanning confocal microscope.
Figure 3.4 (Continued)
XIC and across the entire genome, and, furthermore, our HOP probe discovery pipeline determined that ~40% of the SNPs occurred in genomic sequences suitable to serve as an Oligopaint FISH probe. This density of variants allowed us to design 129-specific and CAST-specific sets of HOP probes targeting the XIC region, each of which consisted of 1,659 oligos. We also designed 9,058 “interstitial” probes that target the same 2.6 Mb XIC region but avoid all SNPs and HOPs and thus should bind both 129 and CAST chromosomes equally well. All three probe sets also avoided the genomic region from which Xist is transcribed, thus giving us the option to perform simultaneous RNA/DNA FISH (Beliveau et al., 2012) by including a fourth probe set consisting of 96 oligos targeting the Xist RNA.

We first simultaneously hybridized AlexaFluor 488 labeled 129 HOP (green), ATTO565 labeled CAST HOP (magenta), and ATTO633 labeled interstitial probes (white) to the three aforementioned MEF lines. As expected, the interstitial probes produced strong staining in all three lines (Supplementary Figure A2.14). A notably different, homolog-specific staining pattern was observed with the HOP probe sets (Figure 3.4B). Specifically, the signals of each HOP co-localized with approximately half of the interstitial probe signals in hybrid EY.T4 129xCAST MEFs (49.5% and 50.5% of interstitial probe signals co-localized with 129 and CAST HOP signals, respectively; n = 111 nuclei, 440 signals; Supplementary Figure A2.14), 100% of the 129 HOP signals co-localized with the interstitial probe signals in 129 MEFs (n = 111 nuclei, 401 signals), and 100% of the CAST HOP signals co-localized with the interstitial probe signals in CAST MEFs (n = 111 nuclei, 452 signals). The homolog-specific staining was highly efficient, with 100% of nuclei displaying signals in all three cell types. It was also robust to differences in the relative concentrations of the two HOPs (Supplementary Figure A2.15) but likely dependent on competition between the HOPs, as low levels of co-staining were occasionally observed (see Supplementary Figure A2.14B, bottom panel, center row), and the addition of
either HOP alone resulted in the HOP signal co-localizing with 100% of the interstitial signals in 129xCAST MEFs \((n \geq 57\) nuclei, 190 signals in both cases; **Supplementary Figure A2.15**).

We then confirmed the specificity of HOPs by taking advantage of the fact that the EY.T4 129xCAST MEF line, which is female, has a pattern of X-inactivation in which the \(X_{\text{CAST}}\) is always the active X chromosome (Xa), and the \(X_{129}\) is always the inactivate X chromosome (Xi) (Yildirim, *et al.*, 2011). Because of this pattern, the \(X_{129}\) is expected to be coated in *cis* with the Xist RNA (Jeon *et al.*, 2012) and thus presents an independent means by which to visually identify the *in situ* position of the \(X_{129}\) chromosome. Accordingly, we performed simultaneous RNA/DNA FISH by using the XIC HOPs in conjunction with an Oligopaint probe set targeting the Xist RNA and observed the tight co-localization of 100% of Xist signals \((n = 101\) nuclei, 183 signals) with signals of the 129 HOP (**Figure 3.4C,D** and **Supplementary Figure A2.16**). In contrast, the Xist signal rarely co-localized with the CAST HOP (6.5% of 183 Xist signals) and only did so when a 129 HOP signal was also co-localized at the same nuclear position, providing strong evidence that our HOPs system can efficiently and reliably distinguish the maternal and paternal homologous chromosomes in the MEF cell culture.

We have also had success with HOPs in Drosophila. Here, we examined F1 hybrids produced from a cross of the 057 and 461 lines from the DGRP (Mackay *et al.*, 2012) and targeted a 4.2-megabase region (89E-93C) that is adjacent to the BX-C on the right arm of chromosome 3. This strategy allowed us to use the BX-C Oligopaint probe set in lieu of a set of interstitial probes to confirm that our HOPs were localizing properly to their genomic targets (**Figure 3.4E**). Comparing the 89E-93C regions of the 057 and 461 lines, we found \(~7\) SNPs per kilobase, which is somewhat higher than the genome-wide average of \(~5\) SNPs per kilobase. We then used our HOP probe discovery pipeline to determine that \(~40\)% of the SNPs occurred in
sequences suitable to serve as Oligopaint FISH probes, of which we selected 6,236 to design a pair of 057-specific and 461-specific HOP probe sets. Excitingly, simultaneous hybridization of the AlexaFluor 488 labeled 057 HOP (green), ATTO565 labeled 461 HOP (magenta), and ATTO633 labeled BX-C (blue) probe sets on spread, polytenized chromosomes isolated from the salivary glands of 057/461 hybrid larvae produced a striking pattern of staining in which two swaths of chromosome, both flanked by a blue BX-C signal, were painted either green or magenta (Figure 3.4E). This pattern of homolog-specific staining was not observed in polytene chromosomes isolated from the homozygous parental lines (Supplementary Figure A2.17). Applying the probes to ovaries, we also found that HOPs are effective in whole-mount tissues (Supplementary Figure A2.18).

Just as the X-inactivation pattern of the EY.T4 cell line offered an independent visual assessment of the reliability of HOPs in mammals, the phenomenon of somatic homolog pairing – the physical pairing of maternal and paternal chromosomes observed in Dipteran insects, such as Drosophila (McKee, 2004) – provided a means by which to test the effectiveness of HOPs in Drosophila. Traditionally, the state of pairing of a given locus is assayed via FISH, wherein paired homologous loci are predicted to produce a single FISH signal, while unpaired loci are predicted to produce two spatially separated signals. However, if HOPs can reliably distinguish homologous loci in situ, we would instead expect two signals in both situations, with the HOP signals being co-localized in the paired state and spatially separated in the unpaired state. To test this idea, we simultaneously hybridized our BX-C probe set (white) and our 057-specific (green) and 461-specific (magenta) HOPs targeting the flanking 89E-93C region to Drosophila embryos that were 6-8 hours old, when homolog pairing is being established (Fung et al., 1998). We observed that the levels of pairing at the BX-C (32% one signal, 68% two signals, 0% no signal, n = 101; Figure 3.4F,G) and the adjacent 89E-93C region (34% co-
localized signals, 66% spatially separated signals, 0% no signal, \( n = 101; \) Figure 3.4F,G) were not statistically different (Fisher’s two-tailed exact \( P = 0.88; \) Figure 3.4G). Importantly, we found the pairing status of these two loci to be highly concordant in individual cells (92.1% concordance with 28.7% both paired and 63.4% both unpaired, Fisher’s two-tailed exact \( P = 6.4 \times 10^{-17}, n = 101 \) nuclei from 2 embryos; Figure 3.4H). These results demonstrate that HOPs provide a reliable readout of the individual behaviors of the paternal and maternal homologs.

Chapter 3 Discussion

In sum, we have presented two advances – Oligopaints enabled single-molecule localization super-resolution imaging of unique genomic regions and HOPs, both of which take advantage of the fully programmable nature of our Oligopaint FISH probes. Together, these tools should enable allele-specific studies of the relationship between gene expression and chromosome organization ranging from overall chromosome positioning to fine scale chromatin structure, including intra- and inter-chromosomal interactions. Given the precision at which we have localized single-molecules in situ, we further anticipate that our technologies will permit the visualization of very short genomic regions, such as those on the scale of enhancers and promoters, with a minimum number of oligo probes. Here, studies may benefit from our capacity to engineer Oligopaint oligos to carry a precise number of fluorophores or binding sites for secondary oligos, thus simplifying the interpretation of fluorescent signals. Our strategies could also be enhanced through the use of multiple STORM activator-reporter dye pairings (Bates et al., 2007), facilitated by secondary oligos, or a highly multiplexed version of DNA-PAINT, called Exchange-PAINT (Jungmann et al., 2014). Finally, we note that since HOPs can produce signals using only one SNP every 1-2 kilobases, they should be generally applicable, including
in humans, where the maternal and paternal genomes differ on average by at least ~1 SNP per kilobase (International HapMap Consortium, 2005; International HapMap Consortium, 2007). As such, a combination of HOPs and Oligopaint-facilitated STORM or DNA-PAINT should enable very high resolution, homolog-specific imaging of chromatin structure, with the potential of companion interstitial probes providing even finer-grain information.

Chapter 3 Materials and Methods

Oligonucleotide libraries

The 27E7-28D3, 89D-89E/BXC, 89B-89D, 19q13.11-q13.12, and 19q13.2-q13.31 libraries were synthesized by MYcroarray (Ann Arbor, MI). The 19q13.32-q13.33, HoxB, XIC interstitial, XIC HOPs, and 057/461 HOPs libraries were synthesized by CustomArray (Bothell, WA). The Xist RNA library was synthesized by Integrated DNA Technologies (IDT) (Coralville, IA). Please see Supplementary Table A2.2 for a list of Oligopaint probe sets used in this work.

PCR primers and secondary oligos

Fluorophore-labeled PCR primers, 5’ phosphorylated PCR primers used in the lambda exonuclease protocol, DNA secondary oligos, and 359 satellite probe oligos were purchased from IDT and purified by IDT using HPLC. Unlabeled, unphosphorylated primers were also purchased from IDT and purified by IDT using standard desalting. Fluorophore-labeled LNA/DNA mixers were synthesized by Exiqon (Vedbaek, Denmark) and purified by Exiqon using HPLC. Please see Supplementary Table A2.3 for a list of PCR primer pairs and Supplementary Table A2.4 for a list of secondary oligos used.
**Emulsion PCR amplification of oligonucleotide libraries**

Raw, multiplexed libraries purchased from CustomArray (see above) were amplified using a pair of universal primers using a previously described emulsion PCR protocol (Beliveau et al., 2012) in order to generate template to use in subsequent PCR reactions. Briefly, 100 µl of aqueous PCR master mix was beaten into a 600 µl of 95.95% mineral oil (Sigma M5904):4% ABIL EM90 (Degussa):0.05% Triton X-100 (Sigma T8787) oil phase (v/v/v) at 1000 rpm for 10 minutes at 4°C. Reactions were amplified with the following cycle: 95°C for 2 min; 30 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 20 s, with a final extension step at 72°C for 5 min. After cycling, the DNA was recovered by a series of organic extractions using diethyl ether (Sigma 296082) and ethyl acetate (Sigma 494518), followed by a phenol-chloroform extraction. For detailed emulsion PCR and emulsion breaking protocols, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints).

**Oligopaint probe synthesis**

Oligopaints probes containing secondary oligo binding sites were synthesized as described previously (Beliveau et al., 2012; Beliveau et al., 2014), or via the lambda exonuclease method (see below). In either case, the secondary oligo binding sites were added to Oligopaint probe sets through the use of the following “touch-up” PCR cycle: 95°C for 5 min; 3 cycles of 95°C for 30 s, 60°C for 45 s, and 72°C for 30 s; 40 cycles of 95°C for 30 s, 68°C for 1 min, and 72°C for 30 s, with a final extension step at 72°C for 5 min. If the probe was produced using the ‘two PCR’ method (Supplementary Figure A2.2), the template generated via “touch-up” PCR was further amplified with the following cycle: 95°C for 5 min; 40-43 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 15 s, with a final extension step at 72°C for 5 min. The Xist RNA probe was first extended from 70 bases to 84 bases in a “touch-up” PCR as described previously (Beliveau et al., 2012) prior to one round of labeling PCR using the “touch-up” cycle described above. One
hundred pmol of each primer and 20 pg of template were used per 100 µl of PCR. For detailed probe synthesis protocols, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints).

‘One-day’ probe synthesis using lambda exonuclease

Oligopaint probe sets were amplified using the “two PCR” method described above, but with the unlabeled primer being phosphorylated on its 5’ end. The PCR reaction was then collected, concentrated using spin columns (Zymo D4031), and digested with lambda exonuclease (New England Biolabs M0262). Five units of lambda exonuclease were added per every 100 µl of unconcentrated PCR reaction (e.g. use 50 units if the labeling PCR had a volume of 1 ml prior to concentration by the spin column) and the reaction was incubated at 37°C for 30 minutes in a programmable thermocycler and then stopped by incubation at 75°C for 10 minutes. Finally, the digestion products were concentrated using ethanol precipitation and quantified using spectrophotometry. For a detailed protocol, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints).

Probe design

The 19q13.11-q13.12 library was described previously (Beliveau et al., 2012). The 27E7-28D3, 19q13.2-q13.31, and 19q13.32-q13.33 libraries were constructed from our public database of 32mer probe sequences (Beliveau et al., 2012) (also see http://genetics.med.harvard.edu/oligopaints). The 89D-89E/BXC and 89B-89D libraries consist of 42mer sequences discovered by OligoArray2.1 (Roulliard et al., 2003) run with the following settings: -n 30 -l 42 -L 42 -D 1000 -t 85 -T 99 -s 70 -x 70 -p 35 -P 80 -m "GGGG;CCCC;TTTT;AAAA" -g 44. The XIC Interstitial and Xist RNA libraries consist of 42mer sequences discovered by OligoArray2.1 run with the following settings: -n 30 -l 42 -L 42 -D 1000 -t 85 -T 99 -s 70 -x 70 -p 35 -P 80 -m "GGGG;CCCC;TTTT;AAAA" -g 44.
The XIC HOPs libraries were discovered using OligoArray2.1 settings identical to those used for the XIC Interstitial and Xist RNA libraries, except “-n” was set to 1. The 057/461 HOPs were discovered using OligoArray2.1 settings identical to those used for the XIC HOPs except that “-t” was set to 80. Also see Supplementary Note A2.1.

Construction of SV-40 T-antigen transformed CAST and 129 MEF lines

To generate the CAST and 129 cell lines, primary mouse embryonic fibroblasts (MEFs) were prepared from F1 embryos collected at embryonic day 13.5 from mice of either pure *M. musculus* (129S1/SvImJ) or *M. castaneus* (CAST/EiJ) backgrounds. MEFs were later immortalized by SV-40 T antigen (Brown et al., 1986) and subcloned by limiting dilution to obtain independent clones. The chromosome content of each subclone was screened by DNA FISH using probes against several autosomal genes.

Cell culture

Drosophila clone8 (DGRC 151) and S2R+ (DGRC 150) cells were obtained from the Drosophila Genomics Resource Center. S2R+ cells were grown in serum-supplemented (10%) Schneider’s S2 medium (serum SAFC 12103C; media Gibco 21720) at 25°C. Clone8 cells were grown in M3 medium (Sigma S3652) supplemented with serum (2%; SAFC 12103C), fly extract (2.5%), and 5 µg/ml insulin at 25°C. WI-38 cells (ATCC CCL-75) cells were grown at 37°C + 5% CO2 in serum-supplemented (10%) Dulbecco’s Modified Eagle Medium (DMEM) (serum Gibco 10437; media Gibco 10564). Primary C57BL/6 (GlobalStem), 129, CAST, and EY.T4 129xCAST mouse embryonic fibroblasts were grown in DMEM (Gibco 10313) supplemented with serum (15%, Gibco 10437) and GlutaMAX (Gibco 35050) at 37°C + 5% CO2. Penicillin and streptomycin
(Gibco 15070) were also added to both insect and mammalian cell culture media to final concentrations of 50 units/ml and 50 µg/ml, respectively.

**Preparation of sample slides for FISH**

Slides containing fixed insect and mammalian tissue culture cells were prepared for FISH as described previously (Beliveau *et al.*, 2012; Beliveau *et al.*, 2014). Briefly, 100 µl of a 1x105-1x106 cells/ml suspension in rich media was spotted onto a poly-L-lysine coated slide and allowed to adhere for 1-3 hours in tissue culture conditions (e.g. 37°C, 5% CO2 for mammalian cells). Slides were then washed in 1X PBS, fixed in 1X PBS + 4% (w/v) paraformaldehyde for 10 minutes, rinsed in 1X PBS, washed in 2X SSCT, washed in 2X SSCT + 50% formamide (v/v), and finally transferred to 2X SSCT + 50% formamide for storage at 4°C until use. For a detailed protocol, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints). For STORM imaging, samples were prepared in the same way except that 22x30 mm #1.5 coverslips were used in place of microscope slides. For DNA-PAINT imaging, samples were prepared in the same way except that Lab-Tek II 8 chamber coverglass vessels (Nunc) were used in place of microscope slides and no poly-L-lysine was used.

**Two-color co-localization FISH**

FISH was performed as described previously (Beliveau *et al.*, 2012; Beliveau *et al.*, 2014), with the 20-30 pmol of secondary probe simply being added to a 25 µl hybridization mix in parallel with 30 pmol of primary probe. Briefly, slides were warmed to room temperature (RT), incubated for 2.5 minutes in 2X SSCT + 50% formamide at 92°C, then incubated for 20 minutes in 2X SSCT + 50% formamide at 60°C. A hybridization cocktail consisting of 2X SSCT, 50% formamide, 10% (w/v) dextran sulfate, 10 mg of RNase (Fermentas EN0531), and Oligopaint
probes was then added to the cells and sealed beneath a 22x22 mm #1.5 coverslip using rubber cement. Slides were denatured for 2.5 minutes at 92°C on the top of a water-immersed heat block and allowed to hybridize overnight at 42°C in a humidified chamber. The next day, slides were washed for 15 minutes in 2X SSCT at 60°C, then for 10 minutes in 2X SSCT at RT, then for 10 minutes in 0.2X SSC at RT. Slides were then mounted in SlowFade Gold + DAPI (Invitrogen S36938) under a 22x30 mm #1.5 coverslip and sealed with nail polish. For a detailed FISH protocol, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints). In the instance where the secondary probe was added sequentially, the primary hybridization was performed as described above, except that the secondary probe was not included in the hybridization mix and the second and third wash steps were both shortened to 5 minutes. After these washes, 30 pmol of secondary probe was added in 25 µl of 2X SSCT and sealed under a 22x30mm #1.5 coverslip with rubber cement, then allowed to hybridize for the times indicated in Supplementary Figure A2.6 at 60°C on the top of a water-immersed heat block. The slides were then washed for 10 minutes in 2X SSCT at 60°C, then for 5 minutes in 2X SSCT at room temperature (RT), then for 5 minutes in 0.2X SSC at RT, and finally mounted as described above.

3D FISH for STORM

A protocol adapted from a previously reported simultaneous RNA FISH/3D DNA FISH protocol (Beliveau et al., 2012) was used for STORM. Sample coverslips were warmed to RT, then rinsed in 1X PBT (1X PBS + 0.1% v/v Tween-20). Coverslips were then incubated in an aqueous 1 mg/ml NaBH₄ solution for 7 minutes, then rinsed 5 times in 1X PBT. Coverslips were then incubated in 1X PBS + 0.5% (v/v) Triton X-100 for 10 minutes, then rinsed in 1X PBT. Coverslips were then incubated for 30 minutes in 1X PBS + 20% (v/v) glycerol, and then flash-frozen by immersion into liquid nitrogen. Coverslips were allowed to thaw, placed back in 1X
PBS + 30% glycerol, then flash-frozen again. This process was then repeated one additional
time (3 total flash-freezes). Coverslips were then rinsed in 1X PBT, then incubated in 0.1N HCl
for 5 minutes, and then rinsed twice in 2X SSCT. Coverslips were then incubated in 2X SSCT +
50% formamide (v/v) for 5 minutes, then incubated in 2X SSCT + 50% formamide at 60°C for 20
minutes. At this point, 30 pmol of primary probe and 40 pmol of secondary probe were added to
25 µl of the hybridization cocktail described for ‘Two-color co-localization FISH’ and the
coverslips were sealed to glass slides using rubber cement (the glass slide acts as a ‘coverslip’
in this instance). Samples were denatured for 2.5 minutes at 78°C on the top of a water-
immersed heat block and allowed to hybridize overnight at 47°C in a humidified chamber. The
next day, coverslips were washed as described for ‘Two-color co-localization FISH’ and stored
in 1X PBS at 4°C prior to mounting in STORM imaging buffer (see below) immediately prior to
imaging. For a detailed protocol, please see the Oligopaints website
(http://genetics.med.harvard.edu/oligopaints).

3D FISH for DNA-PAINT imaging

FISH was performed as described for ‘3D FISH for STORM,’ except that the 1X PBS + glycerol
and liquid nitrogen steps were omitted, and instead of being mounted in SlowFade Gold + DAPI
samples were instead transferred to 1X PBS supplemented with 500 mM NaCl and 10 nM
ATTO655 labeled imager strands (Jungmann et al., 2010; Jungmann et al., 2014).

XIC HOPs 3D FISH and simultaneous RNA/3D DNA FISH with HOPs

3D FISH was performed using a streamlined version of a previously reported simultaneous RNA
FISH/3D DNA FISH protocol (Beliveau et al., 2012). Briefly, slides were warmed to RT, rinsed in
1X PBS, then rinsed in 1X PBT. Slides were then incubated for 15 minutes in 1X PBS + 0.5%
(v/v) Triton X-100, then rinsed in 1X PBT. Slides were then incubated for 5 minutes in 0.1N HCl,
then rinsed three times in 2X SSCT. Slides were then incubated in 2X SSCT + 50% formamide (v/v) for 5 minutes, then incubated in 2X SSCT + 50% formamide at 60°C for 60 minutes. At this point, 40 pmol each of primary probe (129 – AlexaFluor 488 label; CAST – ATTO565 label; XIC Interstitial and Xist RNA – ATTO633 label) and 50 pmol each of secondary probe (129 – 2X AlexaFluor 488 labeled Secondary 5; CAST – 2X ATTO565 labeled Secondary 1; XIC Interstitial and Xist RNA – 2X ATTO633 labeled Secondary 6) were added to 25 µl of the hybridization cocktail described for ‘Two-color co-localization FISH.’ If RNA FISH was being performed, RNase was omitted from the hybridization cocktail. Slides were denatured for 3 minutes at 78°C on the top of a water-immersed heat block and allowed to hybridize overnight at 47°C. The next day, slides were washed and mounted as described for ‘Two-color co-localization FISH.’ For a detailed protocol, please see the Oligopaints website (http://genetics.med.harvard.edu/oligopaints).

**HOPs FISH on Drosophila salivary polytene chromosomes**

A protocol from (Cai et al., 2010) was used for the dissection and preparation of chromosome squashes from Drosophila salivary glands. FISH was then performed as described for ‘Two-color co-localization FISH,’ with 20 pmol of primary Oligopaint probe set and secondary oligo being added per reaction for each probe used. Secondary oligos dual-labeled with AlexaFluor488, ATTO565, and ATTO633 were used with the 057 HOP, 461 HOP, and BX-C probe set, respectively.

**Hybridization to whole-mount Drosophila ovarioles:**

A protocol modified from (Dernburg et al., 1996) was used. Females of the genotype y1#8 (wild-type) were aged 24–48 h and then the ovaries were dissected in 1X PBS. Briefly, the dissected ovaries were fixed in a cacodylate fixative buffer (McKim et al., 2009) for 4 minutes. During the
fixation, the ovaries were teased apart toward the germarium tip. After the fixative was removed, the ovaries were transferred from the dissecting dish to a 0.5 ml Eppendorf tube and washed four times in 2X SSCT. The ovaries were then gradually exchanged into 2X SSCT + 50% formamide (v/v) with a series of 10 minute washes in 2X SSCT + 20% formamide, then in 2X SSCT + 40% formamide, and then two washes in 2X SSCT + 50% formamide. The ovaries were then predenatured in 2X SSCT + 50% formamide and heated to 37°C for 4 hours, 92°C for 3 minutes, and finally 60 °C for 20 minutes. Ovaries were then allowed to settle and the 2X SSCT + 50% formamide was removed prior to the addition of 36 μl of hybridization solution [2X SSCT + 50% formamide + 10% (wt/vol) dextran sulfate] and 200 pmol each of primary Oligopaint probe sets suspended in a total volume ≤4 μl of ddH₂O. The tissue and solution were gently mixed by flicking the tube and then heated to 91°C in a thermal cycler for 3 minutes, followed by incubation overnight at 37°C in the dark. Following the overnight incubation with primary probes, 2X SSCT + 50% formamide was added to the sample and washed for 30 minutes at 37 °C. Supernatant was removed and 200 pmol of each secondary oligo was then added in ~50μl of 2X SSCT + 50% formamide at 37 °C for 30 minutes. Following this incubation, two consecutive washes in 2X SSCT + 50% formamide were done at 37°C, followed by one 10 minute wash in 2X SSCT + 20% formamide and four rinses in 2X SSCT, all at room temperature. After settling, excess 2X SSCT was removed and the ovarioles were mounted in SlowFade Gold + DAPI (Invitrogen S36938).

**HOPs FISH in whole-mount Drosophila embryos**

We collected embryos from overnight collections on apple juice plates. After collection, we dechonorated the embryos by submerging them in 50% bleach for 90 seconds, followed by a thorough wash in ddH₂O. For fixation, embryos were placed in PBS containing 4% (w/v) formaldehyde, 0.5% (v/v) Nonidet P-40, and 50 mM EGTA, plus 500 μl Heptane for 30 minutes.
The aqueous phase was removed and replaced with 500 µl MeOH and mixed vigorously for 2 minutes. The embryos were allowed to settle and were washed two times in 100% MeOH and stored for up to a week at -20°C. Prior to FISH, the embryos were rehydrated in 2X SSCT. FISH were then performed as described above for ovarirole.

**Widefield and confocal microscopy and image processing**

Slides were imaged using an Olympus IX-83 widefield epifluorescent microscope using a 60X oil NA 1.42 lens and an Olympus XM-10 camera or a Zeiss LSM-780 laser scanning confocal microscope using a 63x oil NA 1.40 lens. Olympus images were captured and analyzed using Olympus CellSens software, and Zeiss images were captured and analyzed using Zeiss Zen software. Images were processed using the respective microscope-specific software and Adobe Photoshop.

**Quantification of FISH signals**

FISH signals were counted manually using Z-stacks (i.e. not using maximum Z projections). Two signals separated by an edge-to-edge distance of < 1 µm were considered a single focus. The staining efficiency for a given channel (% Labeling) indicates the number of nuclei with at least one focus in a given experiment. In two-color experiments, % Co-localization indicates the percentage of signals produced by the secondary oligo that also had a co-localizing signal from the primary probe.

**Computational modeling of STORM localizations on polymer structures**

Polymers were simulated as follows. We first generated a random walk on a 3D lattice by adding monomers at random to open lattice points next to the growing end of a chain. Steps in each Cartesian direction were selected with equal probability, subject to the constraint that an...
accepted position be unoccupied by existing monomers. Growing chains that got stuck (more than 10 rejected moves) had their the terminal 10 monomers erased and were restarted growing. After assembling this initial random walk for the desired number of monomers, we used the Bond Fluctuation Method (Carmesin and Kremer, 1988) and Pivot Algorithm (Lal, 1969; Madras and Sokal, 1988) to equilibrate the polymer. Polymer chains were converted to STORM images by assigning to each monomer a random number of switching cycles, drawn from an exponential distribution as observed for switching of Cy5 (Dempsey et al., 2011). A small number of background localizations with uniform spatial distribution were then added to the position list. Gaussian white noise was added to the position of each localization to account for limited localization precision. These final “dye” positions were rendered as STORM images in an identical fashion to that used for our raw dye localization data following spot fitting. To simulate the effect of reduced localizations, a random subset of the total localizations was removed prior to rendering. Parameters used: Number of monomers = 600 or 1500, mean number of localizations = 2, sigma for localization precision localization = 1 monomer diameter.

**STORM microscopy**

STORM images were taken on a customized Olympus IX-71 inverted microscope configured for high angle oblique incidence excitation with a 647nm laser and 100x 1.43 NA oil immersion objective. Microscope construction was previously described (Dempsey et al., 2011). Imaging buffer was used as previously described for single color imaging with AlexaFluor 647 (Dempsey et al., 2011), using 1% (v/v) 2-mercaptoethanol as a thiol. Samples were selected in an experimenter-blind fashion and imaged at 60Hz for 32,000-65,000 frames (based on molecule localization rate). Photo-activation of dyes was tuned with a 405 laser for which the intensity was increased steadily throughout the image acquisition in order to maintain an approximately
uniform molecule localization rate for the first half of the acquisition. The same rate of 405 amplification was used for all cells imaged within a sample.

**STORM image construction**

Molecule localization movies were fit using the 3D-DAOSTORM algorithm (Babcock *et al.*, 2012). Localizations were plotted as single points or as Gaussian spots with widths normalized to the number of photons measured per localization using custom software written in MATLAB® (see https://github.com/ZhuangLab/matlab-storm).

**DNA-PAINT microscopy**

Fluorescence imaging was carried out on an inverted Nikon Eclipse Ti microscope (Nikon Instruments) with the Perfect Focus System, applying an objective-type TIRF configuration using a Nikon TIRF illuminator with an oil-immersion objective (CFI Apo TIRF 100×, NA 1.49, Oil) yielding a pixel size of 160 nm. Two lasers were used for excitation: 488 nm (200 mW nominal, Coherent Sapphire) and 647 nm (300 mW nominal, MBP Communications). The laser beam was passed through cleanup filters (ZT488/10 and ZET640/20, Chroma Technology) and coupled into the microscope objective using a multi-band beam splitter (ZT488rdc/ZT561rdc/ZT640rdc, Chroma Technology). Fluorescence light was spectrally filtered with emission filters (ET525/50m and ET700/75m, Chroma Technology) and imaged on an EMCCD camera (iXon X3 DU-897, Andor Technologies). Images were acquired with a CCD readout bandwidth of 3 MHz at 14 bit, 5.1 pre-amp gain and no electron-multiplying gain using the center 256x256 px of the CCD chip. Imaging was performed using HILO illumination (Tokunaga *et al.*, 2008). The laser power density for DNA-PAINT imaging was 283 W/cm2 at 647 nm. A total of 15,000 frames at a frame rate of 10 Hz were collected, resulting in ~25 min imaging time.
DNA-PAINT image construction

Super-resolution DNA-PAINT images were reconstructed using spot-finding and 2D-Gaussian fitting algorithms implemented in LabVIEW (Jungmann et al., 2010; Jungmann et al., 2014). Localizations are represented Gaussian spots with widths normalized to the localization accuracy. A simplified version of the DNA-PAINT software is available for download at http://www.dna-paint.net or http://molecular-systems.net/software.
Chapter 4:

Conclusion
Continuing to expand the utility of Oligopaints

The work presented in Chapters 2 and 3 describes a flexible platform for generating biinformatically designed, oligo-based, ssDNA FISH probes. As technologies such as massively parallel oligo production, microscopy, and fluorophore synthesis continue to improve, we anticipate that the number of technical and biological questions that can be interrogated by Oligopaints will continue to expand. Moreover, many of these new questions may be the result of entirely novel applications of our probe design and synthesis pipelines in addition to further improvements in the resolution and scalability of our original method. Indeed, several ongoing projects in the Wu laboratory as well as some in the laboratories of our collaborators are already pushing the technology into entirely new directions. As there are many potential areas in which Oligopaints could be improved and many new applications that could be explored, a narrower subset that is of particular interest to the author are presented below.

Targeting smaller genomic intervals with Oligopaints

The continued identification of small functional elements such as enhancers, promoters, and small RNAs has led to an ever-increasing demand to visualize genomic features on the scale of a kilobase and smaller (ENCODE Project Consortium, 2004; Kellis et al., 2014). Targets of this size are particularly challenging for Oligopaints. Indeed, as a maximum of ~20-30 of our 32mer or 42mer oligos can fit into a window of this size, only a small number of primary oligos can be used. Probe sets of such low complexities are predicted to be much more sensitive to issues such as template skewing or loss during amplification, as the loss of even a single oligo from the pool would reduce the maximum number of fluorophores recruited to the target locus by five percent or more (e.g. 19/20 vs. 20/20 if the target were fully saturated with a probe set
composed of 20 singly-labeled oligos); in contrast, the loss of a single oligo from probe sets with complexities in the thousands or tens of thousands would only reduce the maximum number of fluorophores recruited by a fraction of a percent. One way to potentially mitigate this issue is to use column-synthesized oligos as the source of Oligopaint probe molecules instead of oligo pools created by massively parallel array synthesis, as column-based synthesis typically produces much greater quantities of each oligo (nanomole – micromole scale, vs. attomole – femtomole scale for array-based synthesis) (reviewed in Kosuri and Church, 2014). Thus, for very low complexity probe sets, large quantities of each probe molecule can be synthesized individually and pooled together to form the final probe set. Moreover, since FISH is typically performed with ~10-100 picomole of Oligopaint probe per assay (Beliveau et al., 2012; Beliveau et al., 2014), pools of column-synthesized oligos do not need to be amplified by PCR or other methods and can be used directly as ssDNA probes.

Low complexity probe sets targeting very short genomic regions may also benefit from fluorescent signal amplification strategies, as a technique that uniformly and specifically increases fluorescent signal at the site of each hybridized oligo would place less reliance upon each individual oligo in the probe set. This amplification strategy could be as simple as the secondary oligo strategy and its related branched LNA adaptation (Figure 3.1B, Supplementary Figure A2.7-A2.9), which only require the addition of one or a few extra oligos to the hybridization reaction. This strategy could be further enhanced with the use of ‘next-generation’ fluorescent labels such as quantum dots instead of organic fluorophores, as quantum dots are 10-100 fold brighter (Figure 4.1A) and are resistant to photobleaching (Bentolila, 2010; Yusuf et al., 2011; Chen O et al., 2013; Caroline Kim, Yue Chen, Brian Beliveau, José Hernandez, Mouni Bawendi, and Chao-ting Wu, unpublished data; reviewed in Michalet et al., 2005; Ioannou and Griffin, 2010). Alternatively, more complicated signal
Figure 4.1 Signal amplification strategies for small target detection with Oligopaints. (A) A graph showing the brightness of commercial quantum dots relative to ‘bright’ organic fluorophores in the green, red, and far-red portions of the visible spectrum. Brightness was determined by multiplying the molar extinction coefficient ($\varepsilon$) of each fluorophore or quantum dot by its quantum yield (Q.Y.). The top labels present the fold difference in brightness of the quantum dot relative to its paired organic fluorophore. As quantum dots do not have absorbance maximums, 405 nm was chosen as the wavelength for $\varepsilon$. $\varepsilon$ and Q.Y. values for the Alexa Fluor and Qdot ITK products were obtained from Life Technologies; values for ATTO 565 were obtained from atto-tec. (B) A schematic of the HCR strategy. The tethered signal amplification structure is programmed to only form in the presence of an initiator sequence, which could be provided by a ‘MainStreet’ 5’ overhang on an Oligopaint probe molecule. Image is reproduced with permission from (Choi et al., 2010). (C) A schematic of the rolling circle amplification strategy. In this depiction, a closed circle of ssDNA is created by the ligation of a ‘padlock’ probe. Oligopaint probe molecules could also be programmed to have MainStreet overhangs that directed the formation of this type of structure (Sonny Nguyen and Chao-ting Wu, unpublished data). Image is reproduced with permission from (Larsson et al., 2010).
amplification strategies could be employed, such as the hybridization chain reaction (HCR) method developed in the laboratory of Niles Pierce (Dirks and Pierce, 2004; Choi et al., 2010, Choi et al., 2014). In a HCR, the presence of an ‘initiator’ sequence triggers the isothermal assembly of a polymerized signal amplification structure from fluorescently labeled oligos that are programmed to primarily adopt a closed hairpin configuration in solution (Figure 4.1B). Signal amplification could also be potentially achieved via enzymatic methods such as rolling circle amplification (Nilsson et al., 1994; Nilsson, 2006; Larsson et al., 2004; Larsson et al., 2010; Sonny Nguyen and Chao-ting Wu, unpublished data), wherein a polymerase with strand-displacement activity and a circular ssDNA template are used to generate many copies of the template in situ that can subsequently be detected by hybridization with a labeled oligo complementary to the amplified ssDNA (Figure 4.1C). Future work will be required to determine if any of these strategies combine well with Oligopaints.

Nanoscale imaging of chromosomes with SMLM and Oligopaints

While the strategies outlined above have the potential to increase the sensitivity of Oligopaints, even very small regions will still appear as diffraction-limited signals when imaged using conventional microscopy. Thus, it may remain difficult to draw conclusions about the degree of co-localization or physical interaction between two small loci, especially if they are located in close proximity on the linear sequence of the chromosome. SMLM offers a powerful alternative in this setting, as its resolution is currently an order of magnitude lower (~10-20 nm vs. 200+ nm), with the potential to drop even further as scientific cameras become more sensitive and organic fluorophores become brighter. The prospect of imaging very small genomic regions with SMLM is particularly appealing when considering the high multiplexability of DNA-PAINT: with the Exchange-PAINT method developed by the laboratory of Peng Yin (Jungmann et al., 2014),
hundreds if not thousands or more ‘colors’ can be visualized sequentially in the same fixed sample. In this strategy, single molecule fluorescence events are created by the transient hybridization of short, fluorescently labeled ‘imager’ oligos to a complementary target sequence present in the ‘MainStreet’ of Oligopaint probe molecules (Figure 3.3A). The Yin lab has engineered dozens of imager sequences, with each being designed to be orthogonal to both the genome being targeted and all other imager sequences; thus, many different ‘colors’ can be achieved by performing sequential imaging reactions with distinct imager strands (Figure 4.2). If the number of orthogonal imager sequences was even further increased, one might be able to follow the trajectory of an entire interphase chromosome through the nucleus in sub-kilobase increments, which would yield an unprecedented sequence-specific view of local chromosome organization on the nanometer scale. This type of approach would be aided by emerging light sheet technologies and related methods that facilitate the creation of high signal:noise focal planes micrometers or more away from the surface of the imaging objective (Voie et al., 1993; Abrahamsson et al., 2013; Chen et al., 2014; reviewed in Lim et al., 2014), as current methods remain limited in the axial depth in which single-molecule fluorescence events can be reliably detected (reviewed in Lim et al., 2014).

**Quantitative imaging of chromosome structure with SMLM and Oligopaints**

In addition to offering very high spatial resolution, SMLM techniques also provide abundant quantitative information, as each single-molecule localization is considered a discrete, countable event. When paired with FISH, the frequency and spatial distribution of single-molecule fluorescence events detected by SMLM have tremendous potential to provide quantitative measurements of chromatin structure, both in terms of the three-dimensional structure of the
target locus and variations in its labeling density. The latter may serve as a very good corollary for fine-scale chromatin compaction changes. For instance, the number of localizations detected per FISH focus or a unit area or volume could provide a quantitative measure of compaction. Such a system would be particularly powerful if applied at sites of allelic non-equivalence, such as those observed in random X-inactivation (reviewed in Jeon et al., 2012), parent-of-origin imprinting (reviewed in Bartolomei and Ferguson-Smith, 2011), and random monoallelic expression (reviewed in Chess, 2012) in combination with the HOPs technology and an ‘interstitial’ probe, with the latter serving as the substrate for SMLM.

Figure 4.2 The Exchange-PAINT method. (A) Schematic illustrating how multiple cycles of sequential imaging can be performed with multiple imager strands (P1, P2, ..., Pn). (B) Diagram of the DNA origami used to test the Exchange-PAINT method illustrating the scale of the origami used. Note that 70 nm is much smaller than the diffraction limit (~250 nm). (C) Example image of ten distinct origami structures produced from ten sequential imaging cycles of the same field, demonstrating that multiplexed detection ability of Exchange-PAINT. Scale bar: 250 nm. (D) Zoomed, pseudocolored views of each origami structure. Notice the lack of cross-talk between ‘colors.’ Scale bar: 25 nm. Figure is adapted with permission from (Jungmann et al., 2014).
Moving beyond FISH

While the work presented here has focused solely on FISH, we believe the platform we have created could readily be applied to other types of experiment. Ultimately, we have provided researchers with a simple and effective bioinformatic method to identify short, genomically unique sequences with tunable thermodynamic properties and a robust set of protocols for generating large quantities of the corresponding single-stranded or double-stranded DNA or RNA oligos from the products of massively parallel array synthesis. These oligos can be labeled with a broad spectrum of molecules in essentially any position, and can also be extended, digested, or truncated by a variety of enzymatic methods. Thus, given a final desired structure, there are likely many molecular biological avenues that can be taken using our platform to produce a pool of oligos with the correct attributes. We hope our work can be extended to facilitate the rapid and cost-effective generation of reagents for experiments such as the solution phase capture of nucleic acid associated protein complexes using biotinylated oligos (Déjardin and Kingston, 2009; Simon et al., 2011; Caroline Kim, Brian Beliveau, and Chao-ting Wu, unpublished data), the generation of complexes pools of guide RNAs for targeted live cell imaging using the Cas9/CRISPR system (Chen B et al., 2013; Andrews Agbleke, Caroline Kim, and Chao-ting Wu, unpublished data), the production of substrates for fluorescence polarization assays to assess the binding of macromolecules to sequences of interest (reviewed in Moerke, 2009), or the large-scale production of antisense oligos for the parallel knockdown of many transcripts (reviewed in Martinez et al., 2013; Hirose and Mannen, 2015). Indeed, it would be of great satisfaction to the author if the pipelines described here were co-opted for a myriad of diverse purposes.
References


Boveri, T.H. *Ergebnisse über die Konstitution der chromatischen Substanz des Zelkerns* (Fischer, Jena 1904).


Gustafsson, M.G. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* 198, 82-7 (2000).


Patwardhan, A. Subpixel position measurement using 1D, 2D and 3D centroid algorithms with emphasis on applications in confocal microscopy. *J. Microsc.* 186, 246-257 (1997).


Appendix 1:

Supplementary material related to Chapter 2
**Figure A1.1** The density of probe sequences in two probe sets. Probe density (# probe sequences per kb) is plotted in 100-kb windows vs. chromosomal position for four adjacent 20,020 oligo probe sets targeting Xq13.1-q21.1 in the human genome (A) and three probe sets composed of 25,000 (green), 2 x 25,000 (blue), and 2 x 52,500 (red) oligos targeting 41E3-60D14 on Drosophila 2R (B).
Figure A1.2 The number of 1-kb intervals inputted into OligoArray (in thousands) for which 0 – 22 probes were found in the C. elegans (ce6), D. melanogaster (dm3), A. thaliana (tair10), M. musculus (mm9) and human (hg19) genomes.
Figure A1.3 Box-whisker plots showing the distance in bp between each probe and its nearest neighbor for each chromosome of each genome mined. The most extreme 5% of each data set was omitted in order to scale the Y-axis such that the remaining 95% were clearly visible. The red line represents the median, while the box encompasses the second and third quartiles (25% - 75%), and the whiskers bound the first and fourth quartiles. Note that the Drosophila genome assembly puts heterochromatin in separate contigs: “2LH” corresponds to the heterochromatin of the left arm of chromosome 2, “3RH” corresponds to the heterochromatin of the right arm of chromosome 3, etc. “U” represents repetitive sequences that could not unambiguously be mapped to a single chromosome.
The Oligopaints .bed format

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Probe Sequence</th>
<th>TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr2</td>
<td>90958302</td>
<td>90958333</td>
<td>GTCCCCCTCTTGGGCTGCTCTGGCATGCCTGGG</td>
<td>89.14</td>
</tr>
<tr>
<td>chr2</td>
<td>90959330</td>
<td>90959361</td>
<td>CAACAAGGCAAGGCTAATCGTGGAGGTGACCCG</td>
<td>86.00</td>
</tr>
<tr>
<td>chr2</td>
<td>90959388</td>
<td>90959419</td>
<td>GAGACGGAGCAGGCTACCGTGAAGGTTGCGGTTG</td>
<td>89.92</td>
</tr>
<tr>
<td>chr2</td>
<td>90959446</td>
<td>90959477</td>
<td>CCAATGACAAATGTTGGCGACAGACAGGCCG</td>
<td>83.61</td>
</tr>
<tr>
<td>chr2</td>
<td>90959491</td>
<td>90959522</td>
<td>TGACAGTGCGGATGGCAAGGGCGGTGACCCG</td>
<td>88.31</td>
</tr>
<tr>
<td>chr2</td>
<td>90959533</td>
<td>90959564</td>
<td>CCGTACATTAGAGCCTCCTCAAGGGTCAAGTTT</td>
<td>83.14</td>
</tr>
<tr>
<td>chr2</td>
<td>90960142</td>
<td>90960173</td>
<td>AGAGAGGAGAGGAAATCTGAGGTGGCTCCAGAG</td>
<td>81.85</td>
</tr>
</tbody>
</table>

Python scripts for working with Oligopaints .bed files

<table>
<thead>
<tr>
<th>Script Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>grabLines.py</td>
<td>Returns a specific range of lines from a .bed file. E.g. the first 10,000 lines in the file.</td>
</tr>
<tr>
<td>grabRegion.py</td>
<td>Returns all the probes from a .bed file falling within the specified chromosomal region. E.g. all of the probes within a chromosomal band or gene.</td>
</tr>
<tr>
<td>orderFile.py</td>
<td>Appends primer sequences to Oligopaints probes using user-specified primers and banding patterns.</td>
</tr>
<tr>
<td>probeRegion.py</td>
<td>Tiles a specified number of probes across a region of interest. E.g. 1,000 probes across a gene body.</td>
</tr>
<tr>
<td>probeNumber.py</td>
<td>Returns all possible contiguous sets of probes in a .bed file that meet a user-specified number of probes. E.g. sets of 10,000.</td>
</tr>
<tr>
<td>sortFile.py</td>
<td>Sorts a given file by the specified column in either ascending or descending order.</td>
</tr>
<tr>
<td>windowSize.py</td>
<td>Returns all contiguous probe sets that cover a specified window size in kilobase pairs (e.g. 50 kb) from a .bed file.</td>
</tr>
</tbody>
</table>

Figure A1.4 Working with the database files for probe sequences. (A) An example of the Browser Extendable Data (.bed) format used to detail the genomic locations of probes taken from human chromosome 2. Our .bed files are tab-delimited text files with 5 columns: the chromosome or scaffold name, the genomic coordinates of the start (5') position of each probe, the genomic coordinates of the end (3') position of each probe, the sequence of each probe, and the melting temperature (TM) of each probe as calculated by OligoArray. (B) A list of python scripts for working with Oligopaints .bed files that will be available on the Oligopaints website. The grabLines.py and grabRegion.py scripts allow users to isolate a region of interest in our .bed files, while sort.py, probeRegion.py, probeNumber.py, windowSize.py, and order.py facilitate the design and ordering of probe sets. Detailed instructions for using these scripts are provided on the Oligopaints website.
**Figure A1.5** Oligopaints can detect small targets in human metaphase spreads. (A) A TYE563 (Cy3 mimic; red) labeled 200 oligo probe set targeting 10 kb at human 4p16.1 was hybridized to normal male metaphase spreads (Abbott Molecular). (B) A TYE563 (Cy3 mimic; red) labeled 850 oligo probe set targeting 52 kb also at 4p16.1 was hybridized to normal male metaphase spreads (Abbott). (A) and (B) are single Z slices, and an inset for each shows an enlarged image of the labeled chromosome. DNA was identified with DAPI staining (blue). Scale bars: 10 µm.
Figure A1.6 Oligopaints work under a range of hybridization conditions. (A) A probe set composed of 20,020 TYE563 (Cy3 mimic) labeled oligos targeting 2.1 Mb on Xq13.1 was hybridized at 400 nM to WI-38 (XX) cells using our standard protocol (see Materials and Methods), but at varying hybridization temperatures. Hybridization was conducted overnight (16 hours). For each condition, a representative image is presented, as well as qualitative scores for signal intensity and level of background on a scale from “-” (undetectable) to “+++++”. The qualitative scores were determined by examining the entire slide, not just the representative image, and are only meant to be compared between samples in the same experiment. Images are maximum Z projections and raw (i.e. completely unprocessed). Scale bars: 10 µm. (B) The same as (A), except that all hybridizations were carried out at 37 °C for a variable amount of time.
Figure A1.6 (Continued)

A  Testing hybridization temperature (overnight hybridization)

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Signal</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 °C</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>23 °C</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>37 °C</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>42 °C</td>
<td>++++</td>
<td>+</td>
</tr>
<tr>
<td>50 °C</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>60 °C</td>
<td>++</td>
<td>++++</td>
</tr>
</tbody>
</table>

B  Testing hybridization time (37°C hyb.)

<table>
<thead>
<tr>
<th>Time</th>
<th>Signal</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 min</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td>30 min</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>2 hr</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>4 hr</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>19 hr</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>44 hr</td>
<td>++++</td>
<td>++</td>
</tr>
</tbody>
</table>
Figure A1.7 Oligopaints are compatible with a range of denaturation conditions. A probe set composed of 20,020 TYE563 (Cy3 mimic) labeled oligos targeting 2.1 Mb on Xq13.1 was hybridized at 400 nM to WI-38 (XX) cells using our standard protocol (see Chapter 2 Materials and Methods), except that the pre-hybridization and denaturation conditions were varied. The 2.5 min 92 °C pre-hybridization incubation in 2X SSCT + 50% (vol/vol) formamide was included or omitted as indicated. All samples were incubated at 60°C in 2X SSCT + 50% formamide for 20 min prior to hybridization. Samples were either denatured at 75 °C or 92 °C in a hybridization buffer with a formamide concentration (% F) of 50% or 70% as indicated. In the cases where the denaturation was performed in 70% formamide, the hybridization was carried out in 70% formamide as well; else, the denaturation and hybridization were performed in 50% formamide. For each condition, a representative image is presented, as well as qualitative scores for signal intensity and level of background on a scale from “-” (undetectable) to “+++++”. The qualitative scores were determined by examining the entire slide, not just the representative image, and are only meant to be compared between samples in the same experiment. Images are maximum Z projections and raw (i.e. completely unprocessed). Scale bars: 10 µm.
Figure A1.8 Oligopaints are compatible with a range of wash conditions. A probe set composed of 20,020 TYE563 (Cy3 mimic) labeled oligos targeting 2.1 Mb on Xq13.1 was hybridized at 400 nM to female WI-38 (XX) cells using our standard protocol (see Chapter 2 Materials and Methods), except that the temperature and formamide concentration (% F) of the 15 min post-hybridization wash was varied. For each condition, a representative image is presented, as well as qualitative scores for signal intensity and level of background on a scale from “-” (undetectable) to “+++++”. The qualitative scores were determined by examining the entire slide, not just the representative image, and are only meant to be compared between samples in the same experiment. Images are maximum Z projections and raw (i.e. completely unprocessed). Scale bars: 10 µm.
**Figure A1.9** Repeated rounds of hybridization with Oligopaints lead to specific signals. WI-38 (XX 2N) cells were targeted in sequential hybridizations using a probe set composed of 20,020 TYE563 (Cy3 mimic; red) labeled oligos targeting 2.1 Mb at Xq13.1 (Hyb. #1), a probe set composed of 20,020 TYE665 (Cy5 mimic; green) labeled oligos targeting 2.5 Mb at Xq13.2 (Hyb. #2), and a probe set composed of 20,020 6-FAM (white) labeled oligos targeting 3.0 Mb at Xq13.3-q21.1 (Hyb. #3). Once a slide was hybridized with a probe set labeled with a given fluorophore (e.g. Tye563/Cy3 for Hyb. #1), the slide was then scanned for the presence of that fluorophore after all successive hybridizations. Images are maximum Z projections. Scale bars: 10 µm.
Figure A1.10 Simultaneous RNA/DNA FISH with Oligopaints. (A) A UCSC genome browser screenshot of the probe set targeting DNA and the probe set targeting RNA. The probe set targeting DNA is centered on the X-inactivation center (XIC), spans 2.5 Mb at Xq13.2, and has a gap at the site of the Xist RNA to avoid cross-talk between the two probe sets. The probe set targeting the Xist RNA consists of 96 oligos targeting exon 1. (B) A TYE563 (Cy3 mimic; red) labeled probe set targeting the XIC and a 6-FAM (green) labeled probe set targeting the Xist RNA were simultaneously hybridized to WI-38 (XX 2N) or MRC-5 (XY 2N) cells in the presence or absence of RNase. Representative maximum Z projections are shown. Scale bars: 10 µm. (C) The hybridization efficiencies and hybridization patterns observed. As noted in the Chapter 2 Materials and Methods, foci were counted manually using images captured by the Zeiss LSM-780. CL = co-localized; NCL = not co-localized. Note that no Xist signal is expected from male cells.
Figure A1.11 High-throughput FISH using Oligopaints. (A) A 384-well plate was seeded with Drosophila Kc167 (4N) cells and subjected to our 384-well FISH protocol (Joyce et al., 2012) using a probe set composed of 25,000 TYE563 (Cy3 mimic) labeled oligos targeting a 2.7-Mb region at 50D1-53D7 on chromosome 2R. After automated image capture, images were processed by the same MATLAB focus identification script used to assess chromosome pairing in (Joyce et al., 2012). (B) The MATLAB focus identification script was able to effectively detect FISH signals; as Drosophila pairs its homologous chromosomes, a high percentage of single-focus nuclei are expected.

<table>
<thead>
<tr>
<th>Probe</th>
<th>0 foci</th>
<th>1 focus</th>
<th>2 foci</th>
<th>3 foci</th>
<th>4 foci</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>50D1-53C7</td>
<td>12.9%</td>
<td>78.2%</td>
<td>16.8%</td>
<td>5.0%</td>
<td>0.0%</td>
<td>101</td>
</tr>
</tbody>
</table>
Oligopaints should be amenable to modification and interfacing with other technologies. (A) Libraries could be modified with respect to lengths of primer and genomic sequences, $T_M$, GC%, probe density along the chromosome, and number of primer pairs included. Amplification can be carried out (B) using PCR with or without emulsion or perhaps also (C) by T7 polymerase (Gnirke et al., 2009) or rolling circle (Kool, 1996; Diegelman and Kool, 1998; Zhao et al., 2008; Johne et al., 2009) amplification followed by cleavage. One or more fluorophores could be integrated (D) through the primer, (E) during polymerase elongation, or (F) after amplification via enzymatic action (e.g. terminal transferase) or chemical modification of modified bases brought in by the primer or during elongation. Oligopaints could also carry (G) haptens that could then be visualized by secondary reagents carrying fluorophores, quantum dots (Bentolila, 2010), or possibly (H) electron dense labels rendering the probes amenable for electron microscopy (EM) (Alivisatos et al., 1996; Mirkin et al., 1996; Bell et al., 2012). (I) Both strands can be labeled and used separately, increasing yield per preparation; double-stranded probes may be effective in some circumstances. (J) Single-strandedness can be achieved through alternative strategies, including exonuclease digestion or incorporation of biotin followed by streptavidin-mediated separation. When both strands are labeled, nicking followed by gel purification can be used to separately isolate the two strands. (K) Strategies incorporating fluorophores or other functionalities directly into the genomic sequences would be compatible with elimination of primer sequences. On the other hand, retention of primer sequences (one or both) provides binding sites for (L) secondary oligos, which can carry fluorophores or other functionalities, or (M) branched DNA (Player et al., 2001), or (N) permit rolling circle amplification (Nilsson, 2006; Larsson et al., 2010; figure not meant to imply circular templates remain hybridized to targets.) These are just some examples of potential uses for the primer sequence, which we call Main Street, in recognition of the many functionalities with which it can be populated as well as its capacity to be either single- or double-stranded.
Figure A1.12 (Continued)

A. Generate libraries containing a single or multiple primer pairs.

B. Amplify with or without emulsion.

C. Use as template for T7 or rolling circle generation of short ssDNA or RNA probes.

D. Alter number and position of fluorophores in primers.

E. Incorporate fluorophores during polymerization.

F. Incorporate fluorophores after amplification.

G. Label with haptens.

H. Incorporate labels suitable for EM.

I. Label both strands.

J. Make single-stranded via exonuclease digestion or biotinylation and streptavidin-mediated separation.

K. Remove all primer sequences if label has been incorporated into genomic sequences.

L. Hybridize with secondary oligos.

M. Augment signal with branched DNA.

N. Augment signal using rolling circle amplification.
Table A1.1 32mer probes gave the densest coverage in pilot probe mining experiments. OligoArray was used to mine 500 kb regions for probes of a fixed length of 32, 36, or 40 bases. 36mer probes were only considered for the first 500 kb interval. The total number of probes found for each length is presented for each interval, along with the corresponding probe density in probes/kb. The OligoArray parameters used were: -n 22 -D 1000 -t 75 -T 85 -s 60 -x 60 -p 35 -P 80 –m "GGGGG;CCCCC;TTTTT;AAAAA" -g 52.

<table>
<thead>
<tr>
<th>Region (Mb)</th>
<th>Size Kb</th>
<th># 32mers</th>
<th># 36mers</th>
<th>#40 mers</th>
<th>32mers/kb</th>
<th>36mers/kb</th>
<th>40mers/kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr19: 33.25 - 33.75</td>
<td>500</td>
<td>4067</td>
<td>3286</td>
<td>2544</td>
<td>8.1</td>
<td>6.6</td>
<td>5.1</td>
</tr>
<tr>
<td>Chr19: 33.75 - 34.25</td>
<td>500</td>
<td>4255</td>
<td>-</td>
<td>2435</td>
<td>8.5</td>
<td>-</td>
<td>4.9</td>
</tr>
<tr>
<td>Chr19: 35.25 - 35.75</td>
<td>500</td>
<td>4449</td>
<td>-</td>
<td>2833</td>
<td>8.9</td>
<td>-</td>
<td>5.7</td>
</tr>
<tr>
<td>Chr19: 37.22 – 37.72</td>
<td>500</td>
<td>4001</td>
<td>-</td>
<td>3188</td>
<td>8.0</td>
<td>-</td>
<td>6.4</td>
</tr>
<tr>
<td>Regions combined</td>
<td>2000</td>
<td>16772</td>
<td>N/A</td>
<td>11000</td>
<td>8.4</td>
<td>N/A</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Table A1.2 The hybridization efficiencies of Oligopaints probe sets. For each probe, the labeling %, or the percentage of cells with at least one FISH focus, is given. As noted in the Chapter 2 Materials and Methods, foci were counted manually using images captured by the Zeiss LSM-780. Nuclei were identified by DAPI staining. Each nucleus was assayed in the ZEN software package for the presence of FISH foci by moving through individual X-Y optical sections in the Z plane, rather than using a maximum intensity projection. Overlapping nuclei were excluded from this analysis. Two foci <1 µm apart were considered a single signal. Karyotypes: WI-38 (XX 2N), MRC-5 (XY 2N), Kc167 (XXXX 4N), Drosophila ovary cell types (XX 2N). *Note that Drosophila pairs its homologous chromosomes in somatic cells, so most cells will have only one focus, even in the tetraploid Kc167 line.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Probe set</th>
<th>Cell/tissue</th>
<th>[conc] nM</th>
<th>n</th>
<th>0 foci</th>
<th>1 focus</th>
<th>2 foci</th>
<th>&gt;2 foci</th>
<th>Labeling %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>4p16.1, 10 kb</td>
<td>WI-38</td>
<td>800</td>
<td>137</td>
<td>14</td>
<td>3</td>
<td>120</td>
<td>0</td>
<td>90%</td>
</tr>
<tr>
<td>Human</td>
<td>4p16.1, 52 kb</td>
<td>WI-38</td>
<td>800</td>
<td>126</td>
<td>3</td>
<td>15</td>
<td>108</td>
<td>0</td>
<td>98%</td>
</tr>
<tr>
<td>Human</td>
<td>Xq13.1, 2.1 Mb</td>
<td>WI-38 (XX)</td>
<td>400</td>
<td>119</td>
<td>0</td>
<td>4</td>
<td>115</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Human</td>
<td>Xq13.1, 2.1 Mb</td>
<td>MRC-5 (XY)</td>
<td>400</td>
<td>124</td>
<td>2</td>
<td>117</td>
<td>5</td>
<td>0</td>
<td>98%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>50D1-53C7, 2.7 Mb</td>
<td>Kc167</td>
<td>400</td>
<td>136</td>
<td>7</td>
<td>128</td>
<td>1</td>
<td>0</td>
<td>95%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>50D1-53C7, 2.7 Mb</td>
<td>Oocyte</td>
<td>8000</td>
<td>28</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>100%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>50D1-53C7, 2.7 Mb</td>
<td>Nurse</td>
<td>8000</td>
<td>24</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>100%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>50D1-53C7, 2.7 Mb</td>
<td>Follicle</td>
<td>8000</td>
<td>110</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>100%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>41E3-44C4, 3.1 Mb &amp; 58D2-60D14, 2.6 Mb</td>
<td>Oocyte</td>
<td>8000</td>
<td>28</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>100%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>41E3-44C4, 3.1 Mb &amp; 58D2-60D14, 2.6 Mb</td>
<td>Nurse</td>
<td>8000</td>
<td>24</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>100%</td>
</tr>
<tr>
<td>Drosophila</td>
<td>41E3-44C4, 3.1 Mb &amp; 58D2-60D14, 2.6 Mb</td>
<td>Follicle</td>
<td>8000</td>
<td>110</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>99%</td>
</tr>
</tbody>
</table>
**Table A1.3** Probe sets described in this work. For each probe set, the organism, genome assembly version, chromosome (Chr.), cytological region and span of the region in kilobases (kb), start and stop coordinates for the given genome assembly, complexity (number of oligos in the probe set), and density in terms of probe sequences per kilobase are given. *, two regions targeted with one probe set consisting of 50,000 oligos. †, two regions targeted with one probe set consisting of 105,000 oligos.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Assembly</th>
<th>Chr.</th>
<th>Region</th>
<th>Span kb</th>
<th>Start</th>
<th>Stop</th>
<th>Complexity</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>hg19</td>
<td>4</td>
<td>4p16.1</td>
<td>10</td>
<td>10,187,140</td>
<td>10,197,431</td>
<td>200</td>
<td>19.43</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>4</td>
<td>4p16.1</td>
<td>52</td>
<td>10,168,287</td>
<td>10,219,884</td>
<td>850</td>
<td>16.47</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>X</td>
<td>Xq13.1</td>
<td>2,077</td>
<td>69,659,411</td>
<td>71,736,635</td>
<td>20,020</td>
<td>9.64</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>X</td>
<td>Xq13.2</td>
<td>2,474</td>
<td>71,736,696</td>
<td>74,210,596</td>
<td>20,020</td>
<td>8.09</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>X</td>
<td>Xq13.3 - q21.1</td>
<td>3,030</td>
<td>74,212,582</td>
<td>77,242,917</td>
<td>20,020</td>
<td>6.61</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>X</td>
<td>Xq21.1</td>
<td>2,495</td>
<td>77,243,012</td>
<td>79,738,097</td>
<td>20,020</td>
<td>8.02</td>
</tr>
<tr>
<td>Human</td>
<td>hg19</td>
<td>19</td>
<td>19p13.11 - p13.12</td>
<td>3,006</td>
<td>33,995,100</td>
<td>37,000,740</td>
<td>20,020</td>
<td>6.66</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dm3</td>
<td>2R</td>
<td>41E3 – 44C4</td>
<td>3,089</td>
<td>1,125,804</td>
<td>4,215,062</td>
<td>25,000</td>
<td>8.09</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dm3</td>
<td>2R</td>
<td>44C4 – 50C9†</td>
<td>5,593</td>
<td>4,216,958</td>
<td>9,809,826</td>
<td>52,500</td>
<td>9.39</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dm3</td>
<td>2R</td>
<td>50D1 – 53C7</td>
<td>2,657</td>
<td>9,089,774</td>
<td>12,466,888</td>
<td>25,000</td>
<td>9.41</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dm3</td>
<td>2R</td>
<td>53C9 – 58B6†</td>
<td>5,530</td>
<td>12,466,827</td>
<td>17,996,735</td>
<td>52,500</td>
<td>9.49</td>
</tr>
<tr>
<td>Drosophila</td>
<td>dm3</td>
<td>2R</td>
<td>58D2 – 60D14</td>
<td>2,643</td>
<td>17,996,666</td>
<td>20,639,723</td>
<td>25,000</td>
<td>9.46</td>
</tr>
</tbody>
</table>
Table A1.4 The cost per assay of Oligopaints continues to drop. The middle column represents costs of earlier protocols, which called for the use of emulsion PCR in the generation of probes, and the right-hand column represents costs of current protocols, which do not use emulsion and differ from earlier protocols in terms of template and primer concentrations and ratios. The reagent costs of the middle column are given per bag of 20 mls of emulsion PCR (emPCR) for the KBioscience HC-16 hydrocycler, each of which contains 2.8 mls of PCR master mix. The right-hand column gives reagent costs for a 96-well plate of non-emulsion PCR, corresponding to 10 mls of PCR master mix. The cost of the oligo library is excluded, but all steps from PCR amplification to gel extraction and final precipitation are considered. While our most efficient preparations have produced signal at concentrations as low as 40 nM, we achieve more consistent results in the 400 nM - 800 nM range. Given the much reduced cost per assay achieved by our recent production runs, we recommend non-emulsion PCR for generating FISH probes using labeled primers; emulsion PCR may be useful for preserving stocks of the libraries (Williams et al., 2006).

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Cost per 20 ml emPCR bag (2.8 ml master mix)</th>
<th>Cost per 96-well plate (10 ml master mix)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag DNA polymerase</td>
<td>$8.44</td>
<td>$30.14</td>
</tr>
<tr>
<td>Fluor-labeled primer</td>
<td>$1.74</td>
<td>$12.43</td>
</tr>
<tr>
<td>Unlabeled primer</td>
<td>$0.12</td>
<td>$0.86</td>
</tr>
<tr>
<td>BSA</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>dNTPs</td>
<td>$2.49</td>
<td>$8.89</td>
</tr>
<tr>
<td>Emulsion oil mix</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>20 ml emPCR bag</td>
<td>$10.00</td>
<td>-</td>
</tr>
<tr>
<td>Diethyl ether</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>Ethyl acetate</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>Glycogen</td>
<td>$21.06</td>
<td>$33.60</td>
</tr>
<tr>
<td>Nicking enzyme (Nb.BsrDI)</td>
<td>$15.12</td>
<td>$32.77</td>
</tr>
<tr>
<td>15% TBE-Urea polyacrylamide gel</td>
<td>$3.56</td>
<td>$7.20</td>
</tr>
<tr>
<td>Pipette tips/consumables</td>
<td>Negligible</td>
<td>Negligible</td>
</tr>
<tr>
<td>Total:</td>
<td>$62.53</td>
<td>$125.89</td>
</tr>
<tr>
<td>Average yield/ bag or plate ± SEM</td>
<td>70.6 ± 18.9 pmoles (n=17)</td>
<td>2050 ± 291 pmoles (n=16)</td>
</tr>
</tbody>
</table>

Cost Estimates

- At 20 pmole per 25 μl hybridization (800 nM) $17.86 $1.52
- At 10 pmole per 25 μl hybridization (400 nM) $8.93 $0.76
- At 1 pmole per 25 μl hybridization (40 nM) $0.89 $0.076
Appendix 1, Supplementary Methods

PCR primers

We used the following PCR primer pairs (5' -> 3'). For the Drosophila 44C4-50C9 + 53C9-58B6 probe set, F: GTATCGTGCAAGGGTGAATGC R: TTGATCTCGCTGGATCGTTCT. For the Drosophila 50D1-53C7 and 41E3-44C4 + 58D2-60D14 probe sets, F: ATCCTAGCCCATACGGCAATG R: CATAGAACGGAAGAGCGTGTG. For all human probe sets and the Xist RNA FISH probe set, F: GTATCGTGCAAGGGTGAATGC R: ATCCTAGCCCATACGGCAATG. Primers were synthesized by Integrated DNA Technologies (IDT) with or without a 5’ conjugated fluorophore and purified by IDT using standard desalting or HPLC, respectively.

Cell culture

Drosophila S2 and Kc167 cells (Drosophila Genomics Resource Center) were passaged using standard techniques in serum-supplemented (10%) Schneider’s S2 medium (serum SAFC 12103C; media Gibco 21720) at 25 °C. Mammalian WI-38 (ATCC CCL-75) and MRC-5 (ATCC CCL-171) cells were grown at 37 °C + 5% CO2 in serum-supplemented (10%) Dulbecco’s Modified Eagle Medium (serum Gibco 10437; media Gibco 10564). Insect and mammalian cell culture media was also supplemented with penicillin and streptomycin (Gibco 15070).

DNA precipitation

DNA was precipitated using standard techniques using 1/10th volume 4 M ammonium acetate, 2.25 volumes cold 100% ethanol, and 1/50th volume 20 mg/ml glycogen (Fermentas R0561) as a carrier. Pellets were washed with 2.25 volumes of cold 70% ethanol, and air-dried for 15 min at 42 °C.
**Emulsion PCR amplification**

Emulsion PCR (emPCR) was conducted following a protocol adapted from (1) by Yusuf Murgha (Univ. of Michigan). An aqueous PCR master mix was beaten into a 95.95% mineral oil (Sigma M5904):4% ABIL EM90 (Degussa):0.05% Triton X-100 (Sigma T8787) oil phase (vol/vol/vol) on a controlled stir plate (VWR 947042-646) at 1000 rpm for 15 min at 4 °C, creating a 1:6 water-in-oil emulsion with an average droplet diameter of 8-10 μM. Reactions were cycled in a KBioscience HC-16 hydrocycler using the following cycling conditions: 95°C for 7 min; 35 cycles of 95 °C for 1 min, 60 °C for 2 min, 72 °C for 1.5 min; 72 °C for 12 min.

**Extraction and purification of ssDNA**

Emulsion breaking was conducted following a protocol adapted from (Williams et al., 2006) by Yusuf Murgha (Univ. of Michigan). Emulsions were broken by a series of organic extractions using diethyl ether (Sigma 296082) and ethyl acetate (Sigma 494518). Nucleic acids were then isolated from the recovered aqueous phase by extraction with 25:24:1 Tris-buffered phenol:chloroform:isoamyl alcohol pH 6.7 (Fisher BP-1752).

**Emulsion PCR and emulsion breaking protocols**

For detailed protocols, see the Oligopaints website [http://genetics.med.harvard.edu/oligopaints](http://genetics.med.harvard.edu/oligopaints)

**Nicking endonuclease digestion**

Purified PCR products were concentrated via precipitation, then digested with Nb.BsmI or Nb.BsrDI (New England Biolabs R0706, R0648) for 4 hours at 65 °C following the manufacturer’s protocol. The products of the digestion reactions were subsequently concentrated via DNA precipitation.
**Metaphase FISH protocol**

Slides were denatured for 1.5 min at 70 °C in 2X SSCT + 70% (vol/vol) formamide, and progressively dehydrated with a series of 5 min washes using ice cold 70%, 90% and finally 100% ethanol. Slides were air dried at RT, after which hybridization cocktail was added as described for interphase FISH and allowed to hybridize overnight at 37 °C.

**Gel electrophoresis and gel extraction**

Concentrated digestion products were separated by size using denaturing gel electrophoresis. Samples were first denatured for 5 min at 95 °C in TBE-Urea sample buffer (BioRad 161-0768), then loaded and run for ~30 min at a constant voltage of 200 V on 15% TBE-Urea polyacrylamide gels (BioRad 345-091). Gel slices were then excised and incubated in 600 µl of 0.4 M ammonium overnight at 55 °C at 300 rpm in a shaking incubator. The next day, samples were spun down and the supernatant was transferred to fresh tubes containing glycogen and 100% ice-cold ethanol and concentrated by precipitation.

**Hybridization to whole-mount Drosophila ovarioles:**

A protocol modified from (Dernburg et al., 1996) was used. Females of the genotype $y^{1-8}$ (wild-type) were aged 24–48 h and then the ovaries were dissected in 1X PBS. Following dissection, the ovaries were fixed in a cacodylate fixative buffer (McKim et al., 2009) for 4 min. During the fixation, the ovaries were teased apart toward the gerarium tip. After the fixative was removed, the ovaries were transferred from the dissecting dish to a 0.5 ml Eppendorf tube and washed four times in 2X SSCT. The ovaries were then gradually exchanged into 2X SSCT + 50% formamide (vol/vol) with a series of 10 min washes in 2X SSCT + 20% formamide, then in 2X SSCT + 40% formamide, and then two washes in 2X SSCT + 50% formamide. The ovaries were then pre-denatured in 2X SSCT + 50% formamide and heated to 37 °C for 4 h, 92 °C for 3
min, and finally 60 °C for 20 min. Ovaries were then allowed to settle and the 2X SSCT + 50% formamide was removed prior to the addition of 36 µl of hybridization solution [3X SSC + 50% formamide + 10% (wt/vol) dextran sulfate] and 200 pmol of Oligopaint probe suspended in a volume of up to 4 µl of dH₂O. The tissue and solution were gently mixed by flicking the tube and then heated to 91 °C in a thermal cycler for 3 min followed by incubation overnight at 37 °C in the dark.

Following the overnight incubation, 2X SSCT + 50% formamide was added to the sample and inverted several times to mix thoroughly. Ovaries were allowed to settle and then samples were washed for an hour in fresh 2X SSCT + 50% formamide. Two more 30 min washes in 2X SSCT + 50% formamide were done at 37 °C, followed by one 10 min wash in 2X SSCT + 20% formamide and four rinses in 2X SSCT all at room temperature. Ovarioles were then blocked in 6 mg/ml normal goat serum in 2X SSCT for 4 h at room temperature and then washed three times quickly in 2X SSCT. An anti-C(3)G antibody (1A8-1G2, mouse) used at 1:300 (Page and Hawley, 2001) was added in 2X SSCT and incubated overnight at room temperature. The following day, ovarioles were washed three times in 2XSSCT for 10 min, 1 h, and 1.5 h. Fluorescein-conjugated secondary antibody (Vector, Burlingame, CA, or Jackson Labs, West Grove, PA) was added and incubated for 4 h. Ovarioles were then washed two times quickly in 2X SSCT, once for 3 h, and then overnight at room temperature. After settling, excess 2X SSCT was removed and the ovarioles were mounted in Slowfade mounting media containing DAPI.

Hybridization to Drosophila Polytene Chromosome Squashes:

A protocol from (Cai et al., 2010) was used for the dissection and preparation of chromosome squashes from Drosophila salivary glands. FISH was then performed as in our interphase FISH section above with 20 pmol of each Oligopaint probe used per hybridization.
**Microscopy and image processing**

Slides were imaged using one of three microscopy set-ups: an Olympus IX-81 or IX-83 epifluorescent microscope with a 60X oil NA 1.42 lens using CellSens Dimension software; a Leica SP2 laser scanning confocal microscope with a 63X oil NA 1.40 lens using Leica LCS software; a Zeiss LSM-780 laser scanning microscope with a 63X oil NA 1.40 lens using Zeiss ZEN software. Images were processed in using the respective microscope-specific software and Adobe Photoshop.

**Determining the efficiency of probe hybridization**

Foci were counted manually using images captured by the Zeiss LSM-780. Nuclei were identified by DAPI staining. Each nucleus was assayed in the ZEN software package for the presence of FISH foci by moving through individual X-Y optical sections in the Z plane, rather than using a maximum intensity projection. Overlapping nuclei were excluded from this analysis. Two foci <1 µm apart were considered a single signal.

**Modified 384-well FISH protocol for Oligopaints**

384-well plates were seeded with cells and prepared for FISH as in (Joyce et al., 2012), and FISH was carried out as in (Joyce et al., 2012) with the following modifications: 20 pmol of the 50D1-53C7 Oligopaints probe set was added to each well. Hybridization was conducted overnight at 37 °C in a humidified chamber. Post-hybridization washes included 20 min in 37 °C 2X SSCT and 5 min in RT 2X SSCT. The cells were imaged with an Evotec Opera Confocal Screening Microscope (Perkin-Elmer) with a 63x water immersion lens. Ten fields were imaged per well, with the acquisition of five optical Z-sections through each field.
Simultaneous RNA/DNA FISH

Simultaneous RNA/DNA FISH was performed using a “3D-FISH” protocol adapted from (Lanzuolo et al., 2007). Briefly, WI-38 and MRC-5 cells were incubated in 1X PBS + 0.1% (vol/vol) Tween-20 in the presence or absence of RNase, then permeabilized by incubation in 1X PBS + 0.5% (vol/vol) Triton X-100, 3-4 snap-freezes in liquid nitrogen, incubation in 0.1 M HCl, and incubation at RT in 2X SSC + 50% (vol/vol) formamide. 40 pmol of the probe set targeting the XIC and 20 pmol of the probe set targeting the Xist RNA were then added to each slide as part of the hybridization cocktail described for interphase FISH. Slides were denatured for 3 min at 78 °C, hybridized overnight at 42 °C in a humidified chamber, and washed as described for interphase FISH. For a detailed simultaneous RNA/DNA FISH protocol, see the Oligopaints website.
Appendix 2:

Supplementary material related to Chapter 3
**Figure A2.1** The use of secondary (2°) oligos facilitates multiplexing. A multiplexed Oligopaints library contains multiple probe sets, each carrying its own unique primer sequences. The use of labeled secondary oligos allows every probe set in a multiplexed library to be potentially visualized with the same oligo.
Figure A2.2 Two strategies used to introduce a 5' label and a secondary (2°) oligo binding sequence into Oligopaint probe sets. (A) Cartoons illustrating the two strategies. (B) A list of experimentally tested “back primers” for each secondary oligo binding sequence.

<table>
<thead>
<tr>
<th>2° oligo</th>
<th>Back Primer Sequence 5' -&gt; 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CACCACGCTGCGATAGAAGG</td>
</tr>
<tr>
<td>2</td>
<td>CGCAGCTCCACTTGATCTGC</td>
</tr>
<tr>
<td>3</td>
<td>CGAAGCCAGGTATCCATAGCC</td>
</tr>
<tr>
<td>4</td>
<td>GGTGACCTCGTATACGGC</td>
</tr>
<tr>
<td>5</td>
<td>TAGCGCAGAGGTCCACGG</td>
</tr>
<tr>
<td>6</td>
<td>CACACGCTTCCGTCTTAGG</td>
</tr>
</tbody>
</table>
Figure A2.3 A “one day” Oligopaints synthesis strategy (A) Cartoon illustrating the lambda exonuclease method (B) Representative image of FISH performed using a probe produced via the lambda exonuclease method. 30 pmol of a Cy3-labeled Oligopaint probe set targeting the 27E7-28D3 region of the left arm of Drosophila chromosome 2 (red) was hybridized to spread salivary gland polytene chromosomes (Chapter 3 Materials and Methods). DNA is stained with DAPI (gray). Image is a single Z slice. Note: while we show FISH on a polytene spread here, we have also used probes produced by this method on fly, mouse, and human culture cells and fly tissues and see no qualitative difference between these and probes made via the gel extraction method.
Figure A2.4 Co-localization of primary (1°, Oligopaint) probe sets and secondary (2°) oligos in diploid human WI-38 cells. 30 pmol of Cy3 labeled Oligopaint probe (red) and 30 pmol of 6-FAM labeled secondary oligos (green) were added simultaneously and allowed to hybridize overnight at 42°C. (A) Images from each hybridization, which were captured on a wide-field epifluorescent microscope and are maximum Z projections. (B) Staining patterns observed. n = nuclei scored, see Online Methods for % Labeling and % Co-localization. Similar results were observed in Drosophila clone8 cells (data not shown).
Figure A2.5 Sequential addition of labeled secondary (2°) oligo in diploid human WI-38 cells. 30 pmol of Cy3 labeled primary (1°, Oligopaint) probe set (red) consisting of 20,020 oligos targeting 3.0 megabases at 19q13.11-q13.12 was added and allowed to hybridize overnight at 42°C. Slides were then washed, incubated at 60°C with 30 pmol of 2X Alexa647 2° oligo 1 (green) in 2X SSCT for the indicated amount of time, washed, and mounted (Chapter 3 Materials and Methods). Images and qualitative assessments of the signal and background of the primary and secondary probes on a scale from “-” (undetectable) to “+++++” are given for each experiment. \( n \) = nuclei scored, see Online Methods for % Labeling and % Co-localization. Images were captured on a wide-field epifluorescent microscope and are maximum Z projections.
Figure A2.6 Locked nucleic acids (LNAs) as secondary (2°) oligos. (A) The design used to test LNAs at the 359 satellite on the Drosophila X chromosome. (B) Images of FISH with performed in XY diploid Drosophila clone8 cells. 40 pmol of 2X TYE563 labeled LNA secondary oligo (red) was added in the presence or absence of 20 pmol of a 6-FAM labeled primary oligo (green) and allowed to hybridize overnight at 42°C. Images were captured on a wide-field epifluorescent microscope and are maximum Z projections. (C) Staining patterns observed. \( n \) = nuclei scored, see Chapter 3 Materials and Methods for % Labeling, % Co-localization.
Figure A2.7 Branched LNA (bLNA). (A) Cartoons illustrating the binding of secondary (2°), tertiary (3°), and quaternary (4°) oligos to a primary (1°) oligo probe. For each, the number of fluorophores associated the primary oligo if all binding sites were occupied is given. (B) Sequences of oligos used in bLNA experiments. A “+” sign preceding a base indicates an LNA residue. The Integrated DNA Technologies modification codes '/5TYE563/' and '/3TYE563/' indicate 5' and 3' TYE563 molecules.
Figure A2.8 Branched LNA (bLNA) at the Drosophila 359 satellite. (A) FISH with LNA secondary (2°) oligos in XY diploid Drosophila clone8 cells. 20 pmol of TYE563 labeled primary (1°) probe oligo targeting the 359 satellite on chromosome X (red) was added in the presence or absence of the indicated secondary, tertiary (3°), and quaternary (4°) oligos; 20 pmol of each bLNA oligo and 40 pmol of each 2X TYE563 labeled LNA oligo were added. Hybridizations were performed overnight at 42°C. Images from each are shown, below which are the exposure settings used to acquire TYE563. Images were captured on a wide-field epifluorescent microscope and are maximum Z projections. (B) Staining patterns observed. \( n = \) nuclei scored, see Chapter 3 Materials and Methods for % Labeling.
Figure A2.9 Branched LNA (bLNA) leads to stronger signals. (A) The indicated samples were imaged on a Zeiss LSM-780 laser scanning confocal microscope with a constant gain setting (master gain 700, digital gain 1.5). Representative images are shown as pseudocolored micrographs and with the ‘Range Indicator’ look up table for TYE563. Red pixels indicate saturation. \( n \) = the number of nuclei, % Labeling = percent nuclei with at least one signal, % Saturated = percent nuclei with at least one saturated focus. (B) The distribution of signals for each experiment. \( P \) values were calculated using a two-tailed Fisher’s exact tests. ‘n.s.’ = not significant.
Figure A2.10 Multiple enabled by MainStreet and secondary (2°) oligos. In addition to facilitating multiplexing, secondary oligos could amplify signal by carrying multiple copies of the dye present on the primary (1°) probe. Primary-secondary dye pairings can also be engineered for applications such as multicolor labeling and Förster resonance energy transfer (FRET).
Figure A2.11 Oligopaints primary-secondary pairings and STORM. (A) Images of the Drosophila BX-C produced using the three indicated primary-secondary pairings. Top row: conventional, diffraction-limited images. Bottom row: STORM images of the same foci. (B) Cumulative distribution plot of the number of single-molecule localizations observed per BX-C focus for each of the three primary-secondary pairings. Separate curves in the same color indicate biological replicates, which were combined to produce the medians plotted in Figure 3.2C. n = the number of foci imaged per replicate. (C) Examples of off-target background staining. A single nucleus is shown, with the perimeter approximately outlined. The arrowhead indicates the “on target” cluster of localizations corresponding to the BX-C locus that clearly overlap with the signal in the conventional image. Small independent clusters of localizations are often observed elsewhere in the nucleus (arrows), in addition to stray localizations.
Figure A2.12 The number of single molecule localization events (y-axis) remains relatively constant over time (x-axis) during DNA-PAINT imaging, demonstrating that photo-bleaching does not noticeably reduce the localization rate over the time-scale used for imaging. These data correspond to the two foci presented in Figure 3.3D,E.
**HOP design pipeline**

Obtain the positions of SNPs

Search the reference genome in short windows containing each SNP position for probes with OligoArray

Process OligoArray results and filter unsuitable probe sites

Run snpPopper.py

1. Overlapping probes are collapsed into a single probe

2. The variant bases are written into the probe sequences

Figure A2.13 HOP design pipeline using OligoArray (Rouillard et al., 2003) and snpPopper.py.
Figure A2.14 HOPs targeting the mouse X-inactivation center. (A) Probe sets as in Figure 3.4A. (B) FISH performed in the three indicated MEF lines. Arrows indicate where slight cross-talk between the HOPs was observed. DNA is stained with DAPI (blue). Images were captured on a laser scanning confocal microscope and are maximum Z projections. (C) Staining patterns observed. \( n \) = nuclei scored.
Figure A2.15 Concentration-dependent effects on HOP staining patterns. (A) FISH performed in the hybrid EY.T4 129xCAST line. Each sample received the XIC ‘Interstitial’ probe set (white) and the indicated amount of the 129 (green) and CAST (magenta) XIC HOPs. DNA is stained with DAPI (blue). Images were captured on a laser scanning confocal microscope and are maximum Z projections. (B) Staining patterns observed. n = nuclei scored.
Figure A2.16 RNA/DNA FISH with HOPs. (A) Probe sets as in Figure 3.4A. (B) FISH performed with or without the addition of RNase in the hybrid EY.T4 129xCAST line. DNA is stained with DAPI (blue). Images were captured on a laser scanning confocal microscope and are maximum Z projections. (C) Staining patterns observed. $n$ = nuclei scored.
Figure A2.17 HOPs FISH on spread Drosophila polytene chromosomes. FISH was performed on polytene chromosomes isolated from the salivary glands of 057/057 homozygotes (right), 057/461 hybrids (center), and 461/461 homozygotes (left). In all cases, the BX-C probe set (blue) was hybridized along the 057 (green) and 461 (magenta) 89E-93C HOPs. DNA was stained with DAPI (grey). Bottom panels: zoomed images of boxed regions in the corresponding top panels. Images were captured on a laser scanning confocal microscope and are single Z slices. Note that because Drosophila pairs its homologous chromosomes in somatic cells, only one swath of signal is expected per locus targeted, with separation being occasionally observed (e.g. in the center panels).
Figure A2.18 HOPs FISH on whole-mount Drosophila ovaries. FISH performed with the 057 (green) and 461 (magenta) 89E-93C HOPs. Left: paired polytenized chromosomes in Stage 2 nurse cells. Right: dispersed polytenized chromosomes in Stage 10 nurse cells. DNA is stained with DAPI (blue). Images were captured on a laser scanning confocal microscope and are single Z slices. For a description of the polytene dispersal phenomenon, please see (Dej and Spradling, 1999).
Table A2.1 Secondary (2°) oligos and their binding sequences (i.e. the reverse complement of the secondary oligo sequence). For the locked nucleic acid (LNA) oligos, a “+” sign preceding a base indicates an LNA residue at that position. Melting temperature ($T_m$) values were obtained from Exiqon’s online ‘LNA Oligo $T_m$ Prediction’ tool. $^a$Derived from secondary oligo 5. $^b$Derived from secondary oligo 6. $^c$Derived from secondary oligo 1.

<table>
<thead>
<tr>
<th>2° oligo</th>
<th>Sequence 5’ -&gt; 3’</th>
<th>Binding Sequence 5’ -&gt; 3’</th>
<th>%G+C</th>
<th>$T_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CACACGCTCTTCCGTTCTATGCGACGCTCGGTTG</td>
<td>CACCAGCTGCGCATAAGACACGAGACGCTCGGTTG</td>
<td>59.4</td>
<td>76°C</td>
</tr>
<tr>
<td>2</td>
<td>AGAACGATCCACGCAGAGTCAGGAGCTGCGG</td>
<td>GCACGCGATCCACCTTCAGGATCGCTGCTCCAAGGCTCTTCT</td>
<td>56.2</td>
<td>75°C</td>
</tr>
<tr>
<td>3</td>
<td>CATTGCCGATGGGCTAGGATAGCTGGCTCG</td>
<td>CGAGCCAGGCTGATTGCTAGGGGACTAGGCATACGGCAATG</td>
<td>59.4</td>
<td>75°C</td>
</tr>
<tr>
<td>4</td>
<td>GCATTCCACACCTCAGCGGCAGGAGCCAACGCG</td>
<td>GGTGTCGCCTGGTTAGTCTGGCTGAAGGCTGTA</td>
<td>59.4</td>
<td>76°C</td>
</tr>
<tr>
<td>5</td>
<td>ACACCCTGCCAGCCTCGGAGCCATCTGCGAGCTA</td>
<td>TAGAGCGAGGAGGCACAGCGGTCAAGGGCTG</td>
<td>62.5</td>
<td>79°C</td>
</tr>
<tr>
<td>6</td>
<td>TGATGGCGACACGCGCCAAGCGAGAGCGAGCTCGGTTG</td>
<td>CACACGCTCTCCTCGCTGGCAGGCTGATCA</td>
<td>62.5</td>
<td>78°C</td>
</tr>
<tr>
<td>LNA1$^a$</td>
<td>T+CC+AC+GA+CG+TG+CA+A</td>
<td>TTGAGTCGCTGGA</td>
<td>57.1</td>
<td>79°C</td>
</tr>
<tr>
<td>LNA2$^b$</td>
<td>G+AC+GA+GG+TC+GA+GC+A</td>
<td>TGCTCGACCTGGTC</td>
<td>64.3</td>
<td>80°C</td>
</tr>
<tr>
<td>LNA3$^c$</td>
<td>C+GA+CT+AG+CA+CC+GG+T</td>
<td>ACCGGCTAGTGCG</td>
<td>64.3</td>
<td>82°C</td>
</tr>
</tbody>
</table>
Table A2.2 Oligopaint probe sets used. The genomic homology length (Hom. Len.), organism targeted, genome assembly used, span of the probe set in kilobases, start and stop coordinates of the span, the number of oligos in the probe set (complexity), the number of probes per kilobase of target are given for each probe set.

<table>
<thead>
<tr>
<th>Probe set</th>
<th>Hom. Len.</th>
<th>Organism</th>
<th>Assembly</th>
<th>Chr.</th>
<th>Span kb</th>
<th>Start</th>
<th>Stop</th>
<th>Complexity</th>
<th>Probes/kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>27E7-28D3</td>
<td>32</td>
<td>Drosophila</td>
<td>dm3</td>
<td>2L</td>
<td>680</td>
<td>7,256,488</td>
<td>7,936,487</td>
<td>10,000</td>
<td>14.7</td>
</tr>
<tr>
<td>89B-89D</td>
<td>42</td>
<td>Drosophila</td>
<td>dm3</td>
<td>3R</td>
<td>176</td>
<td>12,281,443</td>
<td>12,457,345</td>
<td>1,641</td>
<td>9.3</td>
</tr>
<tr>
<td>89D-89E/BX-C</td>
<td>42</td>
<td>Drosophila</td>
<td>dm3</td>
<td>3R</td>
<td>316</td>
<td>12,482,502</td>
<td>12,797,965</td>
<td>2,394</td>
<td>7.6</td>
</tr>
<tr>
<td>057 HOP</td>
<td>42</td>
<td>Drosophila</td>
<td>dm3</td>
<td>3R</td>
<td>4,201</td>
<td>12,798,329</td>
<td>16,999,743</td>
<td>6,236</td>
<td>1.5</td>
</tr>
<tr>
<td>461 HOP</td>
<td>42</td>
<td>Drosophila</td>
<td>dm3</td>
<td>3R</td>
<td>4,201</td>
<td>12,798,329</td>
<td>16,999,743</td>
<td>6,236</td>
<td>1.5</td>
</tr>
<tr>
<td>19q13.11-q13.12</td>
<td>32</td>
<td>Human</td>
<td>hg19</td>
<td>19</td>
<td>3,006</td>
<td>33,995,100</td>
<td>37,000,740</td>
<td>20,020</td>
<td>6.7</td>
</tr>
<tr>
<td>19q13.2-q13.31</td>
<td>32</td>
<td>Human</td>
<td>hg19</td>
<td>19</td>
<td>2,126</td>
<td>41,281,436</td>
<td>43,407,336</td>
<td>20,020</td>
<td>9.4</td>
</tr>
<tr>
<td>19q13.32-q13.33</td>
<td>32</td>
<td>Human</td>
<td>hg19</td>
<td>19</td>
<td>2,267</td>
<td>47,707,413</td>
<td>49,974,862</td>
<td>20,020</td>
<td>8.8</td>
</tr>
<tr>
<td>HoxB</td>
<td>42</td>
<td>Mouse</td>
<td>mm9</td>
<td>11</td>
<td>174</td>
<td>96,055,675</td>
<td>96,229,567</td>
<td>1,691</td>
<td>9.7</td>
</tr>
<tr>
<td>XIC Interstitial</td>
<td>42</td>
<td>Mouse</td>
<td>mm9</td>
<td>X</td>
<td>2,562</td>
<td>99,370,122</td>
<td>101,932,522</td>
<td>9,058</td>
<td>3.5</td>
</tr>
<tr>
<td>129 HOP</td>
<td>42</td>
<td>Mouse</td>
<td>mm9</td>
<td>X</td>
<td>2,562</td>
<td>99,370,785</td>
<td>101,933,168</td>
<td>1,659</td>
<td>0.6</td>
</tr>
<tr>
<td>CAST HOP</td>
<td>42</td>
<td>Mouse</td>
<td>mm9</td>
<td>X</td>
<td>2,562</td>
<td>99,370,785</td>
<td>101,933,168</td>
<td>1,659</td>
<td>0.6</td>
</tr>
<tr>
<td>Xist RNA</td>
<td>42</td>
<td>Mouse</td>
<td>mm9</td>
<td>X</td>
<td>9.5</td>
<td>100,669,044</td>
<td>100,678,521</td>
<td>96</td>
<td>10.1</td>
</tr>
</tbody>
</table>
Table A2.3 Primer pairs used. The sequence is given for the forward and reverse primers in each pair, as well as the Oligopaint probe sets that each was used to amplify. Primers that have engineered 3’ nicking endonuclease sites (see Beliveau et al., 2012) have the location of the nicking site underlined.

<table>
<thead>
<tr>
<th>F primer sequence 5’-3’</th>
<th>R primer sequence 5’-3’</th>
<th>Probes used on</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTATCGTGCAAGGGTGAAATGC</td>
<td>GAGCAGTCACAGTCAGAAGG</td>
<td>89D-89E/BXC</td>
</tr>
<tr>
<td>CGCTCGGTCTCCGTTCGTCCTC</td>
<td>GGGCTAGGTACAGGGTTTCAGC</td>
<td>129 HOP, 057 HOP</td>
</tr>
<tr>
<td>CAGGTCGAGCCCTGTAGTACG</td>
<td>CTAGGAGACAGCCCGAGACAC</td>
<td>19q13.32-q13.33, CAST HOP, 461 HOP</td>
</tr>
<tr>
<td>GACTGGTACTCGGTGACTTG</td>
<td>CGTCAGTACAGGGTGATGTC</td>
<td>89B-89D</td>
</tr>
<tr>
<td>GTATCGTGCAAGGGTGAAATGC</td>
<td>ATCTAGCCCATACGCAATAG</td>
<td>19q13.11-q13.12, 19q13.2-q13.31, Xist RNA</td>
</tr>
<tr>
<td>CCAGTGCTCGTGAGAAGTC</td>
<td>CTGCAGAGAAGAGCGAGGTC</td>
<td>HoxB, XIC Interstitial</td>
</tr>
</tbody>
</table>
Table A2.4 Secondary oligos used. For each, the sequence is given, with the position and identity of fluorophores presented using Integrated DNA Technologies modification codes. Note that the secondary oligos contain both a 5’ and 3’ label have an eight-base spacer sequence between the 3’ end of the secondary oligo and the 3’ fluorophore – this is to help minimize quenching if a 5’ fluorophore is present on the primary Oligopaint probe set (see Figure A2.3B). These secondary probes represent older designs in which a short orthogonal sequence was used as the spacer. Our newer secondary probes (e.g. Secondary 1 2X ATTO 565) use a poly-T spacer.

<table>
<thead>
<tr>
<th>Description</th>
<th>Sequence 5’-&gt;3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>2° oligo 1 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ CACACGCTCTTCCGTTCTATGCAGTCGGTG</td>
</tr>
<tr>
<td>2° oligo 2 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ AGAACGATCCAGCGATCAAATGGGAGCTCGG</td>
</tr>
<tr>
<td>2° oligo 3 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ CATTGACGTGATGGGCTAGGTACCTGGCTGCG</td>
</tr>
<tr>
<td>2° oligo 4 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ GCATTCACCTTGACGCTACGGACACC</td>
</tr>
<tr>
<td>2° oligo 5 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ ACACCGCTTGACGTGGACCTCCTGCGCTA</td>
</tr>
<tr>
<td>2° oligo 6 with 5’ label</td>
<td>/56-FAM/ or /5Cy3/ or /5Cy5/ TGATCGACAGGCTCAAGACGGAGACGTGTG</td>
</tr>
<tr>
<td>2° oligo 5 with 3’ Cy5</td>
<td>ACACCGCTTGACGTGGACCTCCTGCGCTA/3Cy5/</td>
</tr>
<tr>
<td>2° oligo 1 2X Alexa647&lt;sup&gt;a&lt;/sup&gt;</td>
<td>/5Alex647N/CACACGCTTTCCGTTCTATGCAGTCGGTGagaatggtg/3AlexF647N/</td>
</tr>
<tr>
<td>2° oligo 1 2X ATTO565</td>
<td>/5ATTO565N/CACACGCTTTCCGTTCTATGCAGTCGGTGagaatggtg/3ATTO565N/</td>
</tr>
<tr>
<td>2° oligo 5 2X Alexa488&lt;sup&gt;a&lt;/sup&gt;</td>
<td>/5Alex488N/ACACCGCTTGACGTGGACCTCCTGCGCTAagaatggtg/3AlexF488N/</td>
</tr>
<tr>
<td>2° oligo 6 2X ATTO633</td>
<td>/5ATTO633N/TGATCGACAGGCTCAAGACGGAGACGTGTGagaatggtg/3ATTO633N/</td>
</tr>
</tbody>
</table>
In our initial report of Oligopaints (Beliveau et al., 2012), we focused solely on oligo probes bearing 32 bases of homology to their genomic target each, as this length gave us high probe coverage in preliminary probe mining runs and was compatible with shorter array formats (e.g. 60mer oligos if 14-base primer sequences are used). We have subsequently explored other probe homology lengths and found that slightly longer genomic homology lengths give higher probe coverage when a higher \( T_M \) threshold is enforced (e.g. minimum \( T_M = 85^\circ C \), vs. \( 75^\circ C \) in our initial report), with the higher \( T_M \) threshold being predicted to allow for more stringent hybridization and wash conditions. While the optimal homology length will likely vary with the \%G+C and repeat content of a given genome and the specific nature of the probe set being designed, we have found 42 bases to be a suitable length for the Drosophila, mouse, and human genomes.
Appendix 3:

Additional graduate school publications
Germline Progenitors Escape the Widespread Phenomenon of Homolog Pairing during Drosophila Development

Eric F. Joyce, Nicholas Apostolopoulos, Brian J. Beliveau, C.-ting Wu

Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Homolog pairing, which plays a critical role in meiosis, poses a potential risk if it occurs in inappropriate tissues or between nonallelic sites, as it can lead to changes in gene expression, chromosome entanglements, and loss-of-heterozygosity due to mitotic recombination. This is particularly true in Drosophila, which supports organism-wide pairing throughout development. Discovered over a century ago, such extensive pairing has led to the perception that germline pairing in the adult gonad is an extension of the pairing established during embryogenesis and, therefore, differs from the mechanism utilized in most species to initiate pairing specifically in the germline. Here, we show that, contrary to long-standing assumptions, Drosophila meiotic pairing in the gonad is not an extension of pairing established during embryogenesis. Instead, we find that homologous chromosomes are unpaired in primordial germ cells from the moment the germline can be distinguished from the soma in the embryo and remain unpaired even in the germline stem cells of the adult gonad. We further establish that pairing originates immediately after the stem cell stage. This pairing occurs well before the initiation of meiosis and, strikingly, continues through the several mitotic divisions preceding meiosis. These discoveries indicate that the spatial organization of the Drosophila genome differs between the germline and the soma from the earliest moments of development and thus argue that homolog pairing in the germline is an active process as versus a passive continuation of pairing established during embryogenesis.

Introduction

During meiosis, the germline nucleus undergoes extensive reorganization to accurately align homologous chromosomes along their entire length, enabling them to recombine and ultimately segregate from one another. Outside of the germline, however, homolog pairing, if it occurs at all, is usually transient and localized to a particular chromosomal region [1–7]. Indeed, the individual somatic chromosomes of many eukaryotes occupy distinct territories in the nucleus [8–10], which would be expected to minimize interactions between homologous chromosomes and thus pairing-mediated changes in gene expression, chromosome entanglements, and loss-of-heterozygosity due to mitotic recombination [11–17]. Consequently, extensive homolog pairing is generally considered a germline-specific phenomenon that is restricted to the early stages of meiosis.

One striking exception is found in Dipteran insects, such as Drosophila, where there is widespread homolog pairing in somatic cells. Such pairing has been documented in embryonic, larval, and adult tissues, with pairing frequencies at individual loci reaching 80% or more [18–23]. These observations have led researchers to speculate that Drosophila represents a major departure from other organisms in terms of nuclear organization. The implications are especially profound with respect to the germline, where it has been widely presumed that the homolog pairing observed during Drosophila meiosis is an extension of the pairing established during embryogenesis [24–30]. Notably, there is evidence for homolog pairing being in place during the mitotic divisions immediately preceding meiosis, consistent with it having been established much earlier in development [27,29,31]. Indeed, such pre-meiotic pairing has been reported to continue uninterrupted into meiosis [27,29,31], which may explain the ability of Drosophila females to maintain interactions associated with meiotic pairing and form the synaptonemal complex (SC) between homologs in the absence of double-strand breaks (DSBs) [32], induction of which is essential for pairing and SC formation in yeast and mammals.

This early pairing in the Drosophila germline is in stark contrast to meiotic pairing in non-Dipteran organisms consisting of distinct soma and germline tissues [30,33–35]; while a recent study showed pairing as early as the final round of pre-meiotic replication in mice, there was no demonstration of pairing earlier to this time point [36]. Here we clarify the origin of germline pairing in Drosophila, refuting a long-held hypothesis that it derives from pairing established during embryogenesis and arguing, instead, for a program of germline pairing that is not initiated until the five mitotic cell cycles just prior to meiosis.


Copyright: © 2013 Joyce et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Ruth L. Kirschstein National Research Service Award to EFJ from NIH/NCI (F32CA157188) and a grant and an NIH Director’s Pioneer Award to CtW from NIH/NIGMS (RO1GM61936 and 5DP1GM106414); the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: twu@genetics.med.harvard.edu
followed by high-resolution microscopy and 3D-image reconstruction by fluorescent in situ hybridization (FISH) in whole-mounted ovaries proceeds [39].

distinguish them from subsequent pre-meiotic stages, we took advantage of an antibody to the cytoplasmic protein SXL, levels of which increase in GSCs and then decrease as differentiation proceeds [39].

Chromosome positioning in individual GSC nuclei was assessed by fluorescent in situ hybridization (FISH) in whole-mounted tissues, using techniques that preserve the nuclear architecture followed by high-resolution microscopy and 3D-image reconstruction. Within each nucleus, a single FISH signal or two signals separated by ≤8.5 μm were considered to represent the paired state of the targeted locus. To evaluate the extent of genome-wide pairing, we used nine FISH probes (Figure 1B and Table S1). Three probes targeted highly repeated sequences of the centromeric heterochromatin, including that of the X chromosome (359), chromosome 2 (AACAC), and chromosome 3 (dodeca). The remaining six probes were generated with Oligopaint technology [40] and targeted single copy euchromatic loci, including two loci (5A and 16E) on the X chromosome, loci on the left (24D) and right (50D) arms of chromosome 2 and the left (69C) and right (100B) arms of chromosome 3. Importantly, these probes were extremely efficient, with 100% of nuclei displaying at least one focus for each probe.

In stark contrast to the assumption that meiotic pairing is an extension of pre-existing somatic pairing and that chromosomes therefore enter the germline in the paired state, we observed extensive separation of homologs in GSCs. Eight out of the nine loci produced two distinct FISH signals in the majority of, if not all, nuclei and were considered unpaired in 75–100% of GSC nuclei (Figure 1C–D) in experiments representing 15–30 ovaries. In fact, the average inter-allelic distances for these loci was equivalent to the radius of the nucleus (2.29 μm; p > 0.05, unpaired t test; Table S1), consistent with a random positioning of the maternal and paternal chromosomes relative to each other. The dramatic deficiency of pairing at these eight loci argues that the genome-wide homolog pairing and subsequent SC formation during meiosis does not derive from a paired state that is extant in the GSC.

The one exception was the X-linked 359 repeat, which was paired in 80% of GSC nuclei (Figure 1D). While this may be indicative of some homolog alignment in GSCs, it may also reflect the proximity of this locus to the rDNA gene cluster, which is spatially confined to the nucleolus, and/or the large size of this repeated region [20,28,41], estimated to be 11 Mb in size [42]. Importantly, two other euchromatic loci proximal (16E) and distal (5A) to the 359 repeat were mostly (75%) and completely (100%), respectively, unpaired, indicating that the X-chromosome is not exceptional in its capacity to pair in GSCs. We have also found the 359 repeat to exhibit atypical pairing dynamics in somatic cells [43].

Homolog pairing is established during the mitotic cell cycles prior to meiosis

The largely unpaired state of GSCs shifted our focus to determining whether, outside of the 359 repeat, pre-meiotic pairing occurs in the female germline to any significant extent. To this end, we looked directly downstream of the GSCs to the differentiating CBs, which number between one and two per germline arm, and, relative to the GSCs, are positioned downstream of the niche. Here, we targeted the euchromatic loci of 5A, 16E, 24D, 50D, 69C, and 100B (Figure 1D and Table S1) and found unambiguous levels (11–35%) of pairing at all but 5A. These findings establish that Drosophila does, indeed, support at least some degree of pairing well before meiosis initiates.

Despite significant levels of euchromatic pairing in CBs, no pairing was detected at the centromeric repeats AACAC and dodeca, suggesting the partial nature of homolog pairing at this stage. In fact, when we performed two-color FISH targeting two loci across a single arm of chromosome 2 in CB nuclei, we did not detect any pairing of the centromeric locus despite 33% (n = 21) pairing of the chromosome arm (Figure 1E).

The partial pairing observed in CBs raised the possibility that complete pairing can be achieved in cells progressing through mitotic divisions prior to the initiation of meiosis. In Drosophila, this program includes four rounds of divisions in which the CB becomes a 2-, 4-, 8-, and ultimately 16-cell cyst of interconnected cystocytes, one of which completes meiosis. Importantly, these stages precede the pachytene stage of meiosis, in which homologous chromosomes are fully synapsed (Figure 2A). In order to evaluate the progression of pre-meiotic pairing, we used the P[acman-GFP] transgene, a transcriptional reporter that is not expressed in GSCs or pachytenes, but is expressed in each of the intervening pre-meiotic stages [39,44], along with an antibody against Spectrin, a cytoskeletal protein that forms a spherical structure called the spectrosome in GSCs and CBs, and an antibody against C(3)G, which identifies the SC in meiotic nuclei [45] (Figure 2A–C). Developing cysts were staged based on the number of RAM-positive cells and by Spectrin staining, which, in cysts, localizes to a branched structure called the fusome. We found that homolog pairing levels rapidly increased through the divisions, with each of six FISH targets reaching maximum levels of pairing (87–95%) by the 8-cell cyst (Figure 2D). Once maximum levels of pairing were achieved, they were maintained throughout the remaining pre-meiotic divisions and into the pachytene stage of meiosis (Figure 2C–D), suggesting that homologous chromo-
somes initiate pairing up to four mitotic divisions prior to meiosis and enter meiosis fully aligned. Importantly, we were able to assess pairing in all four cells of twenty-seven 4-cell cysts, where 87%, 90%, 100%, and 95% of cells showed pairing at dodeca, 24D, 69C, and 100B, respectively. Because the oocyte derives from one cell of a 4-cell cyst, these observations demonstrate that pairing

Figure 1. Homologous chromosomes enter the germline unpaired. A, Left: Schematic of a germarium showing pre-meiotic mitotic cell divisions as well as maturation of the meiotic cysts. The GSCs (purple nuclei) are positioned adjacent to the somatic niche (brown) and express high levels of SXL (green cytoplasm). Each GSC divides asymmetrically to produce a renewed stem cell and a differentiating cystoblast (CB, blue nucleus surrounded by green cytoplasm), which is positioned distal to the niche. The CB will undergo four more rounds of mitotic divisions to form a 16-cell cyst. Following these pre-meiotic stages, the 16-cell cyst will enter meiotic prophase, as defined by the initiation (zygotene) and complete formation (pachytene) of the synaptonemal complex (SC, red) between the paired homologs in two of the sixteen cells. Only a single cell will complete meiosis within each 16-cell cyst to form a mature egg (not shown). Arrow, direction of maturation. Right: Wild-type germarium stained for DNA (blue) and SXL (green). A GSC and CB are indicated by arrows and identified by SXL staining and relative position to somatic niche. Approximately 1–2 GSCs and 1–2 CBs are present in each germarium. Scale bar represents 10 μm. B, Drosophila chromosomes and targets of FISH probes (red). Heterochromatin is denoted in grey and rDNA cluster on the X-chromosome is in purple. C, Image of a GSC nucleus (dashed circle) at the tip of a germarium identified by DAPI (blue) surrounded by cytoplasmic SXL (green) staining and combined with FISH targeting AACAC (red) and dodeca (grey). Two signals for each FISH target represent separated homologous loci. Scale bar represents 5 μm. D, Percentage of nuclei exhibiting paired and unpaired loci in GSCs (left panel) and CBs (right panel). 15–30 ovaries were scored for each stage with a combined total of 242 GSC nuclei and 262 CB nuclei (approximately 30 nuclei for each locus at each stage). E, CB nuclei identified with SXL staining in combination with two-color FISH targeting AACAC (grey) and 24D (red) on Chromosome 2. Cartoon depicts hypothetical arrangement of homologous chromosomes as either unpaired or partially paired. Scale bars represents 5 μm.

doi:10.1371/journal.pgen.1004013.g001
can be observed well before the initiation of meiosis. Analogous data were obtained for the β-cell and 16-cell cysts.

Interestingly, not all chromosome loci achieved pre-meiotic pairing at the same rate; the 359, 314D, and 100B loci reached maximum levels of pairing prior to the 5A, AACAC, dodeca, and 69C loci by one to four divisions (Figure 1D and Figure 2D). This observation is consistent with the higher level of euchromatic as versus centromeric pairing we observed for autosomes in CBs and suggests that, rather than strictly initiating at the centromeres, where SC formation is first observed [46,47], germine pairing may initiate at different rates or times across the genome.

Chromosomes maintain an unpaired state throughout germline development

The results described above refute the long held belief that homolog pairing in Drosophila meiosis is an extension of pairing events established within the embryo and maintained throughout development in all tissues, including the germline. Hence, they encouraged us to assess whether germinal cells ever support somatic levels of homolog pairing during development or, alternatively, whether unpairing represents a global nuclear reorganization specifically in the GSCs. We, therefore, evaluated pairing levels during embryogenesis, as this 24-hour phase of development marks both the onset of somatic homolog pairing as well as the separation of germline and somatic lineages. The germine distinguishes itself~2 hours after egg lay (AEL), with the primordial germ cells (PGCs), from which adult GSCs are derived, forming at the posterior pole of the embryo and becoming identifiable with the germine-specific protein marker VASA [48] (Figure 3A-B). Examination of homologous pairing during embryogenesis has indicated that some sites attain pairing as early as 2 hours AEL [21]. In fact, the ~500 Kb histone locus on chromosome 2 has been reported to pair ~2.5 hours AEL in the soma and PGCs [31], providing reason to believe that PGCs do not differ from somatic cells in their capacity to pair. However, it is unclear if pairing of this locus reflects genome-wide levels or specific features of this locus, such as its transcriptional activity [17,20,21,20,49,50]. Here, we distinguish these alternatives by examining the behavior of four other loci across the genome - two centromeric (AACAC and dodeca) and two single-copy euchromatic loci (24D and 30D).

Confirming our ability to detect the onset of pairing in somatic cells, we examined embryonic 2.5 hours AEL and observed, respectively, 12% and 10% pairing at the centromeric AACAC and dodeca repeats and 22% and 29% pairing at the euchromatic 24D and 50D loci (Figure 3C). In contrast, the PGCs at the posterior pole of the embryos consistently exhibited lower levels of pairing at this stage for all four loci, ranging from 3% to 5% (Figure 3C). We next analyzed embryonic nuclei 14 hours AEL and found 80–100% of somatic nuclei were paired for each of the four loci (Figure 3D), levels consistent with full attainment of somatic homolog pairing [21]. Strikingly, however, homologous chromosomes in the PGCs, which at 14 hours AEL are in contact with somatic cells, remained essentially unpaired, attaining only 0–7% pairing at any of the four loci (p<0.0001; Figure 3D). In these cells, the inter-allelic distances were extensive, averaging 2.5–3.5 μm and, in some cases, reaching as much as 5–6 μm (Figure S1). Additionally, sex-specific differences were not observed, suggesting that, regardless of sex, germinale progenitors do not support genome-wide homolog pairing (Figure S2). Thus, PGCs maintain a predominantly unpaired state of homologous chromosomes throughout embryogenesis. These observations argue that germ cells are never exposed to the widespread pairing observed in somatic cells and thus, represent the only Drosophila tissue identified so far that escapes this phenomenon.

Germline progenitors have large nuclear volumes with chromosomes juxtaposed to the nuclear envelope

To better understand how germline progenitors maintain an unpaired state, we determined whether they might exhibit other distinctive features of nuclear organization. Notably, we observed that the nuclear volumes of PGCs 14 hours AEL (~72.5 μm³) were ~3.3 fold greater than that of neighboring somatic cells (~21.3 μm³, p<0.0001; Figure 4A) and reasoned that larger nuclear volumes could cause increased distances between homolog and thus account for the lower levels of pairing in PGCs. Consistent with this hypothesis, the nuclear volumes of GSCs and CBs (50.5–52.7 μm³) were greater than two times larger than that of the surrounding somatic follicle cells (~21.3 μm³, p<0.0001; Figure 4A). This analysis revealed that, even when inter-allelic distances were normalized, the level of pairing in PGCs remained less than that observed in somatic cells by six to nineteen fold (Figure S3). This outcome suggests that the separation of homologous chromosomes cannot be fully explained by nuclear volume alone.

We next analyzed the global distribution of DNA within the larger GSC nuclei and found a distinct nuclear f-β-diamidinono-2-phenylindole (DAPI) staining pattern compared to somatic follicle cells, indicating a change in chromatin structure. As shown in Figure 4C, surface plots of DAPI fluorescence intensity revealed a non-uniform peripheral staining pattern in GSC nuclei. In contrast, the DAPI fluorescence intensity in adjacent somatic follicle cell nuclei typically displayed a relatively uniform and diffuse staining pattern (Figure 4C). To test whether this DNA distribution could result from the juxtaposition of chromosomes to the nuclear envelope, we used FISH in combination with an antibody against lamin to measure the distance between the nuclear envelope and each of three loci across chromosome 2: the AACAC centromeric repeat and the 24D and 50D loci in the middle of each arm. As predicted, all three loci were equally close to the nuclear envelope in GSCs, with average distances of only 15–20% of the total nuclear radius (Figure 4D). Similar results were found for the nuclei of CB, β-cell cysts, pachytene cells, as well as embryonic PGCs (Figure 4D and Figure S4), indicating that the peripheral localization of chromosomes is adopted early in germline development and maintained into meiosis. In contrast, the same three loci in somatic cells exhibited greater distances from the nuclear envelope, averaging of 25–47% of the radius, with the centromeric locus closest to the nuclear envelope (Figure 4D and Figure S4). This arrangement, in which centromeres are located in the periphery with chromosome arms displaced across the nuclear space, is consistent with the Rabl configuration of chromosomes frequently found in Drosophila somatic cells [51,52]. We conclude that germinal cells may adopt a distinct nuclear structure which, compared to somatic cells, involves placement of chromosomes in close proximity to the nuclear envelope along their entire length.

Conclusion

Our findings reveal extensive separation of homologous chromosomes in germinale progenitors from early embryogenesis until the five mitotic cell cycles just prior to meiosis and, in this regard, align Drosophila with other organisms that establish homolog pairing de novo in the gonad. Importantly, our observa-
Figure 2. Homolog pairing is established during the mitotic cell cycles prior to meiosis. A, Left: Schematic of a gerarium identifying the pre-meiotic stages with BAM (red) and Spectrin (white) and meiotic stages with C(3)G (green). Right: Wild-type gerarium in which GSCs are identified by the position near the niche, absence of BAM staining, and presence of a spectrosome (white). Developing cysts are identified by the presence of BAM staining and a branched fusome (white). DAPI, blue. Approximately 1–2 germline cysts are present in each gerarium, with equal occurrence of the 2-, 4-, 8-, and 16-cell stages. Scale bar represents 10 μm. B, FISH targeting dodeca (grey). 2-cell and 4-cell cysts (the 4th cell is out of focus) identified with BAM:GFP (pseudo-colored red). Scale bars represent 10 μm. C, FISH targeting 24D (red) and AACAC (grey) in a gerarium identifying pachytene nuclei in meiosis with an antibody against the SC protein C(3)G (green). Scale bar represents 10 μm in upper panel and 5 μm in lower panel. D, Percentage of nuclei exhibiting paired and unpaired loci in pre-meiotic stages as well as in meiotic pachytene with FISH targeting AACAC, dodeca, 5A, 24D, 69C, and 100B. Pre-meiotic cysts were identified using BAM:GFP and Spectrin. Pachytene nuclei were identified in a separate experiment using an antibody against C(3)G. 15–30 ovaries were scored for each stage with a minimum of 20 nuclei counted for 2- and 4-cell stages, 40 nuclei for the 8-cell stage, and 80 nuclei for the 16-cell stage (*P < 0.01, **P < 0.0001).

doi:10.1371/journal.pgen.1004013.g002
tions are in agreement with Christophorou et al. (also in this issue), who also found a deficiency of homolog pairing in the GSCs of the Drosophila adult female. The lack of pairing in germline progenitors is especially noteworthy, considering the widespread prevalence of pairing in the somatic tissues of Drosophila. Why should an organism ensure homologous chromosomes remain unpaired in germline progenitors, only to allow pairing beyond the stem cell divisions? One possibility is that, since germline progenitors generate the entire cell population responsible for transmitting the genome to subsequent generations, any negative outcome of that pairing could be propagated to a much greater extent as compared to undesired events occurring downstream of the GSCs and thus have a higher probability of multigenerational consequences.

Our discovery of unpaired homologs in germline progenitors also demonstrates that homolog pairing is not an inevitable feature of Drosophila chromosomes and is consistent with studies arguing that pairing is a controlled process reflecting genes that promote pairing as well as those that antagonize it [14,15,22,23,43,53]. Here, we further propose that potentially undesirable homologous interactions are precluded in Drosophila germline progenitors coordinately with, or due to, the separation of the progenitors from the soma in early embryogenesis. Pairing could also be precluded through a localization of chromosomes to the nuclear periphery. Such a configuration could lead to the formation of chromosome territories that separate homologs in the germline, as opposed to configurations that permit or even promote the pairing observed in the soma (Figure 3). Note that our data do not clarify whether the mechanisms that pair homologous chromosomes in somatic cells are distinct from, or similar to, those that eventually pair homologous chromosomes in the pre-meiotic cells. Nevertheless, to the extent that the mechanisms may be different, our findings are consistent with the notion that germline nuclei may suppress or delay the mechanisms that support pairing in the soma, perhaps through nuclear organization, while, in the pre-meiotic cells, simultaneously permit a separate mechanism that promotes pairing. Indeed, Christophorou et al. (also in this issue) show that pre-meiotic pairing is perturbed in the absence of meiosis-specific proteins such as components of the SC, suggesting that the mechanisms of pre-meiotic pairing cannot be entirely similar to that of somatic pairing.

Interestingly, we found that chromosomes maintain their peripheral localization even during the pre-meiotic 8-cell stage when homology is fully aligned and continue to maintain this configuration into meiosis. Whether this localization is a significant aspect of pre-meiotic pairing will be of interest, as chromosome interactions with the nuclear envelope have been proposed to promote meiotic pairing in several organisms [24–37], as well as influence polyteny pairing in Drosophila [58]. Regardless, our observations establish a distinction between the organization of paired chromosomes in pre-meiotic nuclei (peripheral localization) and that in somatic nuclei (internal localization) in Drosophila.
Figure 4. Germline progenitors contain large nuclear volumes with chromosomes juxtaposed to the nuclear envelope. A, PGC nuclei (VASA, red) are larger than surrounding somatic nuclei in embryonic gonads 14 h AEL. DAPI, blue. Dashed circles denote nuclear periphery. Scale bar represents 10 μm. Right: Average nuclear volume of PGCs and surrounding somatic cells ± SEM. B, Average nuclear volume of germline and somatic follicle cells of the adult ovary ± SEM (**p < 0.0001). C, Wild-type germarium stained for DNA (grey) and SXL (green). Shown on right are cross-sections of representative GSC and somatic nuclei with 3D and 2D (insets) surface plots displaying increased peripheral intensity in the nucleus of the GSC and more uniform intensity in the nucleus of the somatic cell. Scale bar represents 10 μm in the image of the germarium and 5 μm in images of...
In closing, we return to the extraordinary degree of pairing that Drosophila and other Dipteran insects support in their soma. If homologous interactions can lead to negative outcomes, why do these organisms permit a near-organismal wide level of such interactions? One explanation is that somatic homolog interactions may, under some circumstances, confer advantages [12,28,34,59–63] and, consistent with this, transient and localized instances of somatic homolog interactions have been documented or at least implicated in a wide variety of organisms, including mammals [1–7,12,63–65]. Indeed, in light of our discovery that the different tissues of Drosophila can have dramatically different levels of pairing it is possible that greater scrutiny of non-Dipteran species will reveal many more instances of somatic pairing and, hence, evidence that somatic pairing is a widespread potential of genomes in general [12,14,43].

Materials and Methods

Drosophila strains

Drosophila stocks and crosses were maintained on a standard medium at 25°C. For wild-type, we used the y1 w1118 strain. To identify the BAM protein that was used to distinguish the pre-meiotic cyst stages, we crossed y1 w1118 to a strain carrying the transgene P(bamP-GFP) [44], a kind gift from Michael Buszczak (The University of Texas Southwestern Medical Center at Dallas).

Generation of FISH probes

Oligo probes for the 359, AACAC, and dodeca heterochromatic repeats [42,66] were synthesized with either a 5’ Cy5 or Tye3 fluorescent dye by Integrated DNA Technologies (IDT). The design of the probe sequences were previously described [43] and are as follows: 359: GGGATCGTTAGCACTGGTAAT-TAGCTG, AACAC: AACACAACACAACACAACACAACAACAACAC, and dodeca: AGGGGACAGTGCGG. Oligo probes were resuspended in 1×TE at 100 μM concentration and stored at −20°C. Euchromatic probes 5A (4E2-5C10), 16E (16B3-17A2), 24D (24D1-24F1), 50D (50D1-53C7), 69C (69A1-69E6), and 100B (100B9-100D1) were designed and generated using the Oligopaint technology [40]. Briefly, a library of 7500 (24D and 100B), 10000 (5A, 16E, and 69C), and 25000 (50D) unique oligos (MYcroarray) were designed for amplification. Each library was amplified using a common 5’ CY3-conjugated forward primer (5’-CGCTCGGTCTCCGTTCGTCTC) and unlabeled reverse primer (5’-GGGCTAGGTACAGGGTTCAGCgcaatg).

Antibodies

The antibodies and dilutions used were mouse anti-SXL (m18, Developmental Studies Hybridoma Bank [DSHB], 1:10), rat anti-VASA (DSHB, 1:300), mouse anti-Spectrin (DSHB 3A9, 1:50), mouse anti-lamin (DSHB ADL84.12, 1:100) and rabbit anti-GFP (Molecular Probes, 1:300). The mouse anti-C(3)G (1A8-1G2, single nuclei. D. Left: GSC nucleus with FISH targeting 24D (red) and lamin staining (nuclear envelope, green). Scale bar represents 5 μm. Right: Average distance between FISH signals and the nuclear envelope (NE) ± SEM, normalized to the nuclear radius, in germline and somatic follicle cells of the adult ovary. Asterisks denote significant differences in the normalized distances between somatic and GSCs (*p<0.05, **p<0.0001). For each data point, a minimum number of 30 nuclei were scored (see Materials and Methods). doi:10.1371/journal.pgen.1004013.g004

Figure 5. Model for germline nuclear organization. Once a germline cell fate is established in early embryogenesis, homologous chromosomes remain unpaired and localize to the nuclear periphery, creating non-overlapping chromosome territories that may block ectopic pairing. This organization is maintained through development and into the adult GSCs. Germline pairing initiates coincident with germline differentiation (time point denoted as ‘D’ during the pre-meiotic mitotic divisions, ultimately leading to complete homolog alignment and the initiation of meiosis and SC formation. In somatic cells, homologous chromosomes instead adopt a configuration that permits, or even promotes, pairing. Such a configuration might be the Rabl organization, which occurs in early embryogenesis and positions centromeres and telomeres at opposite nuclear poles [21].

doi:10.1371/journal.pgen.1004013.g005

Figure 5. Model for germline nuclear organization. Once a germline cell fate is established in early embryogenesis, homologous chromosomes remain unpaired and localize to the nuclear periphery, creating non-overlapping chromosome territories that may block ectopic pairing. This organization is maintained through development and into the adult GSCs. Germline pairing initiates coincident with germline differentiation (time point denoted as ‘D’ during the pre-meiotic mitotic divisions, ultimately leading to complete homolog alignment and the initiation of meiosis and SC formation. In somatic cells, homologous chromosomes instead adopt a configuration that permits, or even promotes, pairing. Such a configuration might be the Rabl organization, which occurs in early embryogenesis and positions centromeres and telomeres at opposite nuclear poles [21].

doi:10.1371/journal.pgen.1004013.g005
formamide, then in 2
the addition of 36
to settle and the 2
settling, excess 2
2
After washing, the ovaries were then gradually exchanged into
room temperature and three quick washes in 2
immunostaining, the embryos were rehydrated in PBT. Immunostaining
by two 20-min washes in PBT and one 10-min wash in PBS.
secondary antibodies at room temperature for 2 hours, followed
min washes in PBT were performed prior to incubation with
antibodies were incubated at 4
washed three times for 5 min in PBT. Prior to immunostaining,
the ovaries were teased apart and blocked by incubating in PBT
plus 1.5% BSA at room temperature for 1 hour. Primary
antibodies were incubated at 4°C overnight in PBT. Three 20-
min washes in PBT were performed prior to incubation with
secondary antibodies at room temperature for 2 hours, followed
by two 20-min washes in PBT and one 10-min wash in PBS.
For FISH, PBS buffer was replaced with 2×SSC/0.3 M
NaCl, 0.01 M NaCitrate, 0.1% Tween-20) by three quick washes. After
washing, the ovaries were then gradually exchanged into
2×SSC/50% formaldehyde and 10-min washes in 2×SSC/20% formaldehyde,
then in 2×SSC/40% formaldehyde, and then two washes in 2×SSC/50% formaldehyde. Ovaries were then allowed
to settle and the 2×SSC/50% formaldehyde was removed prior to
the addition of 36 μl of hybridization solution (2×SSC, 50% formaldehyde, 10% (v/v) dextran sulfate, RNase) and up to 4 μl of
probe. For heterochromatic targets, 100 pmol of probe was added to
the hybridization. For single-copy euchromatin targets, 200–
400 pmol of Oligopaint probes were added to the hybridization.
To preserve the nuclear structure, chromosomes were denatured at 78°C in a thermal cycler for 30 min followed by incubation overnight at 37°C in the dark. Following hybridization, we performed two 30-min washes of 2×SSC/50% formaldehyde at
37°C, followed by a 10-min wash in 2×SSC/20% formaldehyde at
room temperature and three quick washes in 2×SSC. After
settling, excess 2×SSC was removed and the ovaries were mounted in SlowFade mounting media with DAPI (Invitrogen).

Collection and fixation of Drosophila embryos
We collected 25-hour and 14-hour-old embryos for 1 hour on
apple juice plates and then aged them an additional 1.5 or 13 hours, respectively. 2.5 hours after egg lay (AEL) should
capture embryos in the final 10 min of cell cycle 13 and the first
50 min of cell cycle 14. Due to the time spent manipulating embryos during the dechorionation step (see below), most embryos were aged ~5–10 min longer before development was stopped
during fixation. During imaging, embryos were also staged by the
number and position of primordial germ cells, which are separated
from the soma at the pole 2.5 hours AEL and encapsulated within
the embryonic gonad 14 hours AEL (Figure S3B).

After collection, we dechorionated the embryos by submerging them in 50% bleach for 90 seconds, followed by a thorough wash
in ddH2O. For fixation, embryos were placed in PBS containing
4% formaldehyde, 0.5% Nonidet P-40, and 50 mM EGTA, plus
500 μl Heptane for 30 min. The aqueous phase was removed and
replaced with 500 μl MeOH and mixed vigorously for 2 min. The
embryos were allowed to settle and were washed two times in 100% MeOH and stored for up to a week at ~20°C. Prior to
immunostaining, the embryos were rehydrated in PBT. Immuno-
staining and FISH were then performed as described above for
ovaries.

Microscopy and image analysis
All images were collected using a Zeiss LSM780 laser scanning
confocal microscope with a 63×, N.A. 1.40 lens. We imaged whole
germaria by collecting 200 mm optical sections through the entire
tissue at 1024×1024 or 512×512 resolutions with a digital zoom of
3.0. The analysis of the images was performed by both 3D-image
reconstruction and examining one section at a time using the Zeiss
ZEN 2011 software. FISH foci were counted manually within
each nucleus and the distance between the centers of allelic signals
was measured using the Ortho – distance function, which permits
length measurements in 3D space. 100% of nuclei examined in
this study exhibited at least a single FISH signal, indicating high
hybridization efficiency. Therefore, a single signal was consid-
tered two foci with an inter-signal distance of 0 μm. In some
cases where noted, we normalized the inter-signal distances by
the radius of the nucleus. In these cases, p values were
determined by an unpaired t test.

Two homologs were considered paired if the distance between
their focus centers was ≤0.8 μm or FISH produced a single signal.
To determine the significance between paired states, p values were
calculated by a two-tailed Fisher’s exact test.

To image the pre-meiotic 2, 4, 8, and 16-cell cysts, we focused on the
germarium identified by P[banP-GFP] expression, set the upper and lower limits of the scanned region to
capture the entirety of the cysts, and then scored those nuclei that
were fully contained within the scanned region and were also
unambiguously distinguished from other cell types. This strategy
enabled us to score 93–100% of the cells in any chosen cyst.

To image germline cells in the embryo, we focused on that
region identified by VASA expression, set the upper and lower
limits of the scanned region to capture the majority of the germline
cells, and then scored only those nuclei that were fully contained
within the scanned region and were also unambiguously distin-
guished from somatic cells. This strategy enabled us to score
the majority of germline cells in any chosen embryo. The somatic
embryonic cells that were scored were those that were within the
scanned region containing the scored germline cells.

Measuring nuclear volume
Nuclear envelopes were labeled with an anti-lamin antibody.
Nuclear volumes were calculated based on the nuclear diameter
using the equation $V = \frac{4}{3} \pi r^3$.

Measuring distance between nuclear envelope and FISH signals
Nuclear envelopes were labeled with an anti-lamin antibody.
The ZEN software package was then used to measure the shortest
distance between FISH signals and the nuclear envelope in 3D
space. When two FISH signals were present in the nucleus, only
the shortest distance of the two was scored. p values were
determined by an unpaired t test.

Supporting Information
Figure S1 Extensive separation of homologs in primordial germ cells. Relative frequencies of inter-allelic distances in embryonic
PGCs and somatic cells 14 hours AEL based on FISH targeting
AACAC (upper-left), dodeca (upper-right), 24D (lower-left), and
50D (lower-right). The percentage of nuclei exhibiting paired loci
for this data set is presented in Fig. 3. (TIF)
Figure S2 Male and female germline progenitors do not support genome-wide homolog pairing. A, Primordial germ cells (PGCs, VASA red) in female embryos 14 hours AEL were distinguished from male embryos based on their expression of the female-specific cytoplasmic protein SXL (green). B, Percentage of nuclei paired at AACAC (left-most graph) and 24D (right-most graph) in male and female PGCs as compared to somatic cells in embryos 14 hours AEL. No significant difference in pairing levels were observed between the sexes. For each data point, a minimum number of 20 PGCs and 50 somatic nuclei were scored from a total of 6–7 independent embryos (see Materials and Methods). (TI)

Figure S3 Homolog separation in primordial germ cells is not dependent on nuclear size. Distances between allelic signals by FISH (Figure S1) were normalized to the radius of the nucleus to account for the larger nuclear volumes in PGCs as compared to that in somatic cells in embryos 14 h AEL. Despite this normalization, there was 6–7 times less pairing in PGCs than in somatic cells (\(p<0.0001\)). (TI)

References

Figure S4 Chromosomes from primordial germ cells are in close proximity to the nuclear envelope. A, PGC nucleus with FISH targeting 24D (red, arrowheads) and lamina (black) (nuclear envelope, green). Scale bar represents 5 μm. B, Average distance between 24D FISH signals and the nuclear envelope (NE) normalized to the nuclear radius ± SEM in embryonic PGCs as compared to somatic cells 14 hours AEL. (TI)

Table S1 Homolog pairing frequencies in GSCs and CBs. (DOCX)

Acknowledgments
We are grateful to Mohammad Hamman and Chaminne Foneka for technical assistance and members of the Wu lab and Jean-Rene Huysh for critical reading of the manuscript. We also thank Jean-Rene Huysh for sharing unpublished information.

Author Contributions
Conceived and designed the experiments: EFJ NA BJF. Performed the experiments: EFJ. Analyzed the data: EFJ. Contributed reagents/materials/analysis tools: EFJ NA BJF. Wrote the paper: EFJ.

Nuclear Organization in Germline Progenitors

Visualizing genomes with Oligopaint FISH probes

Brian J. Beliveau¹, Nicholas Apostolopoulos², and Chao-ting Wu¹

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts
²Yale School of Medicine, New Haven, Connecticut

Abstract

Oligopaint probes are fluorescently-labeled, single-stranded DNA oligonucleotides that can be used to visualize genomic regions ranging in size from tens of kilobases to many megabases. This unit details how Oligopaint probes can be synthesized using basic molecular biological techniques as well as provides protocols for FISH, 3D-FISH, and sample preparation.

Keywords

Oligopaint; oligonucleotide; complex DNA library; FISH

INTRODUCTION

The availability of inexpensive DNA oligonucleotide (oligo) libraries containing hundreds to hundreds of thousands of unique oligo species has enabled technological advances in areas such as cytogenetics and the study of nuclear biology. In particular, these libraries can serve as a renewable source of probes for fluorescence in situ hybridization (FISH) and allow researchers to precisely define the sequences to be targeted by the probes. This unit details a strategy that applies a few standard molecular biological protocols to complex libraries of ssDNA oligos in order to produce short, single-stranded, highly efficient Oligopaint FISH probes that label >90% of nuclei in fixed tissue culture cells (Beliveau et al., 2012). Probe generation can be accomplished in a matter of days and at a cost that is below that of commercially available bacterial artificial chromosome (BAC)-based probes (Beliveau et al., 2012). Oligopaint probes can target regions ranging in size from tens of kilobases to many megabases, can be made strand-specific, and can be bioinformatically designed to produce customizable patterns, such as multicolor banding.

Basic Protocol 1 presents the strategy for synthesizing Oligopaint probes, while Basic Protocol 2 details a streamlined protocol for using Oligopaint probes to label interphase and mitotic chromosomes in tissue culture cells. Alternate Protocol 2 describes a strategy for...
3D-FISH (Lanzuolo et al., 2007; Cremer et al., 2008), which avoids high-temperature steps in order to better preserve the three-dimensional morphology of nuclei, while Alternate Protocol 3 describes a strategy for labeling metaphase chromosomes. Finally, Support Protocol 2 addresses the preparation of sample slides for use with Basic Protocol 2 and Support Protocol 2.

**BASIC PROTOCOL 1**

**Generation of single-stranded Oligopaint FISH probes from a complex DNA library**—Oligopaint FISH probes are produced by a series of standard molecular biological techniques, including PCR, DNA precipitation, and gel electrophoresis, and the protocol presented below describes the entire process, starting with the incorporation of fluorescent labels during PCR and finishing with the quantification of the purified ssDNA probe. For a discussion of some of the considerations in probe design, see Critical Parameters. This probe synthesis protocol uses the strategy described by Beliveau et al. (2012), in which a nicking endonuclease recognition sequence (e.g. 5’…GCAATG…3’ for Nb.BsrDI) is included in every molecule of the ssDNA library (see Figure 1) to facilitate the isolation of single-stranded probe molecules after the conversion of the library to dsDNA by PCR amplification. Note that while Figure 1 of Beliveau et al. (2012) illustrates the use of a single nicking endonuclease site, it is also possible to include two distinct nicking endonuclease sites (e.g. Nb.BsrDI and Nb.BsmI, Figure 1 of this publication) such that either strand of the dsDNA duplex can be used to generate FISH probe.

**Materials**

- 10X *Taq* DNA polymerase buffer
- 10 mM dNTP mix
- 200 μM fluorophore-labeled “Forward” primer
- 200 μM unlabeled “Reverse” primer
- 20 pg/μl complex DNA library
- 5 U/μl *Taq* DNA polymerase
- Molecular-biology grade water
- 100% ethanol
- 20 mg/ml molecular biology grade glycogen
- 4M ammonium acetate solution in distilled, deionized water (ddH₂O)
- 70% (vol/vol) ethanol solution in ddH₂O
- 10X *Nb.BsrDI* enzyme buffer
- 10 U/μl *Nb.BsrDI* enzyme
- 1X Tris-Borate-EDTA (TBE) buffer (see recipe)
- 15% TBE + 7 M Urea denaturing polyacrylamide gel
- Low molecular weight DNA ladder
- 2X TBE + urea gel loading buffer containing xylene cyanol FF and bromophenol blue
- 10 mg/ml ethidium bromide solution in molecular biology grade water 0.4 M ammonium acetate solution in ddH₂O
- 0.2 ml thin wall strip tubes or thin wall 96-well plate
50 ml conical tube
2.0 ml microcentrifuge tubes
15 ml conical tube
Programmable thermocycler
Benchtop vortexer
Refrigerated centrifuge
Adjustable heat block
Gel box and power supply
Ethidium bromide staining dish
Benchtop orbital shaker
UV box
Heated vortexer or shaking incubator
Spectrophotometer

NOTE: Add 100-fold more template if using raw ssDNA library (i.e. direct from manufacturer, never amplified via PCR) as many of the raw oligos may not be suitable templates for amplification. Raw libraries can be amplified using unlabeled “F” and “R” primers and purified using a PCR-cleanup kit in order to provide a greater amount of template material for labeling PCRs.

1. Set up a PCR master mix using the component ratios listed below and aliquot into 0.2 ml thin wall strip tubes or a thin wall 96-well plate. Do not add more than 100 μl of master mix per tube/well.

1X PCR master mix
10 μl 10X Taq DNA polymerase buffer
2 μl 10 mM dNTP mix
0.5 μl 200 μM fluorophore-labeled “Forward” primer
0.5 μl 200 μM unlabeled “Reverse” primer
1 μl 20 pg/μl complex DNA library
1 μl 5 U/μl Taq DNA polymerase
85 μl molecular biology grade water

The scale of the PCR reaction should correlate with the desired number of FISH assays. In an efficient probe preparation, ≥20% of fluorescently-labeled primer added to the PCR reaction will be recovered as ssDNA FISH probe. Thus, a 10 ml PCR reaction using 10,000 pmol of labeled primer can be expected to produce 2,000 pmol of FISH probe, or 100 20 pmol FISH assays.

NOTE: The proofreading 3’→5’ exonuclease activity of some high-fidelity thermostable DNA polymerases can lead to DNA degradation in subsequent steps. A phenol-chloroform extraction is recommended after the PCR is completed if your polymerase has strong exonuclease activity. This extraction is not necessary for most standard polymerases.
2. Transfer the PCR reactions to a programmable thermocycler and run the following program:

**Labeling PCR program**

1. 95°C 5 minutes
2. 95°C 30 seconds
3. 60°C 30 seconds
4. 72°C 15 seconds
5. Repeat steps 2–4 42X (43 total cycles)
6. 72°C 5 minutes
7. End

This program assumes a primer $T_M$ of ~60°C for both primers; raise or lower the annealing temperature accordingly if necessary.

3. Pool the cycled PCR reactions in a 50 ml conical tube.

To accelerate this step for large volumes, extract the PCR reaction using a multichannel pipette and transfer to a disposable reagent reservoir in a PCR or tissue culture hood.

4. Set up a DNA precipitation in the 50 ml falcon containing the PCR product: for every 600 μl of PCR reaction, add 2 μl 20 mg/ml glycogen, 65 μl 4M ammonium acetate, and 1350 μl ice-cold 100% ethanol. Vortex vigorously and aliquot into 2.0 ml microcentrifuge tubes.

5. Incubate the precipitations at −80°C for 35 minutes or for >2 hours at −20°C. Precipitations can be left at −20°C overnight or indefinitely; this step represents a convenient stopping point.

6. Spin the precipitation at max speed for 1 hour at 4°C in a refrigerated centrifuge. The precipitation can also be spun at room temperature if a refrigerated centrifuge is not available.

7. Carefully aspirate off the 100% ethanol using a vacuum trap or micropipette. A prominent pellet should be visible after the spin. Depending on the fluorophore used, the pellet may appear colored.

8. Add 1350 μl ice-cold 70% to each tube, taking care not to disturb the pellet.

9. Spin the precipitation at max speed for 30 minutes at 4°C in a refrigerated centrifuge.

10. Carefully aspirate off the 70% ethanol using a vacuum trap or micropipette.

11. Air dry the pellets by placing open tubes onto a 42°C heat block for 15 minutes.

12. Add 60 μl of molecular biology grade water to each pellet. Incubate for 30 minutes at 37°C to resuspend. Vortex each tube after the 37°C incubation, then spin briefly in a microcentrifuge. Shaking at 1400 rpm in a heated vortexer will speed up the resuspension. This step represents a convenient stopping point – samples can be left at 4°C for several weeks or frozen at −20°C and left indefinitely.
13. Set up a nicking endonuclease digestion: for every 60 µl of precipitated PCR product, add 7.5 µl 10X nicking enzyme buffer and 7.5 µl 10 U/µl Nb.BsrDI. Aliquot the digestion into 0.2 ml thin wall strip tubes or a thin wall 96-well plate. Do not add more than 25 µl of master mix per tube/well. Incubating the digestion in a single 0.6 ml thin wall tube or 1.7 ml microcentrifuge tube may result in reduced nicking efficiency.

14. Incubate the digestion for 4 hours at 65°C in a programmable thermocycler, then heat-inactivate the enzyme by incubating for 20 minutes at 80°C. Incubating for >4 hours at 65°C is not recommended as it may result in star activity. While Nb.BsrDI and Nb.BsmI require incubation at 65°C, other nicking enzymes may require incubation at 37°C.

15. Pool the digestion reaction in a 15 ml conical tube. Set up a DNA precipitation: for every 600 µl of digestion reaction, add 2 µl 20 mg/ml glycogen, 65 µl 4M ammonium acetate, and 1350 µl ice-cold 100% ethanol. Vortex vigorously and aliquot into 2.0 ml microcentrifuge tubes. Molecular biology grade water can be used to adjust the volume of the pooled digestion reaction to a balanced number of 2.0 ml microcentrifuge tubes (e.g. a multiple of two or three) if desired. The DNA will be very concentrated at this step; a precipitate may form as soon as the digestion is mixed with 100% ethanol.

16. Incubate the precipitations at −80°C for 35 minutes or for >2 hours at −20°C. Precipitations can be left at −20°C overnight or indefinitely; this step represents a convenient stopping point.

17. Spin the precipitation at max speed for 1 hour at 4°C in a refrigerated centrifuge. The precipitation can also be spun at room temperature if a refrigerated centrifuge is not available.

18. Carefully aspirate off the 100% ethanol using a vacuum trap or micropipette. A very large pellet should be visible after the spin. Depending on the fluorophore used, the pellet may appear colored.

19. Add 1350 µl ice-cold 70% to each tube, taking care not to disturb the pellet.

20. Spin the precipitation at max speed for 30 minutes at 4°C in a refrigerated centrifuge.

21. Carefully aspirate off the 70% ethanol using a vacuum trap or micropipette.

22. Air dry the pellets by placing open tubes onto a 42°C heat block for 15 minutes.

23. Add 40 µl of molecular biology grade water to each pellet. Incubate for 30 minutes at 37°C to resuspend. Vortex each tube after the 37°C incubation, then spin briefly in a microcentrifuge. Shaking at 1400 rpm in a heated vortexer will speed up the resuspension. Step 24 can be started during the 37°C incubation.

24. Pre-run a 15% TBE-Urea polyacrylamide gel in pre-warmed (to ~60 °C) 1X TBE at constant wattage for at least 30 minutes. Use a wattage that maintains a buffer temperature of 55°C – 60°C.
It is essential that the gel is sufficiently hot to prevent the renaturation of the nicked DNA. Renaturation will result in the appearance of slow-migrating smears in the gel.

Caution: the gel casing may crack if the temperature is too high.

25. While the gel is pre-running, mix the resuspended precipitation products 1:1 with 2X TBE-urea gel loading buffer and aliquot into 0.2 ml thin wall strip tubes. Load one gel well volume (e.g. 40 μl) per tube. Also mix an aliquot of low molecular weight DNA ladder 1:1 with 2X TBE-urea gel loading buffer and load into a 0.2 ml thin wall strip tube.

26. Denature the samples for 5’ at 95°C in a programmable thermocycler, then transfer directly to ice.

27. Rinse each gel well with 1X TBE (from the gel box) using a 1000 μl micropipette to blast out any urea that may have settled in the well prior to loading the samples.

28. Load the samples and run at the same constant wattage as the pre-run (again maintaining temperature of 55°C – 60°C) until the bromophenol blue marker dye is near the bottom of the gel (~15–20 minutes for an 8.7 cm gel).

The labeled ssDNA product of interest will essentially co-migrate with the xylene cyanol FF marker dye, running slightly slower. Prominent colored bands may be visible depending on the fluorophore used.

29. Carefully remove the gel from its casing and add to a staining dish containing 50 ml of 0.6 μg/ml ethidium bromide solution (3 μl of 10 mg/ml ethidium bromine solution in 50 ml ddH₂O).

30. Stain the gel for 5 minutes with gentle mixing on a benchtop orbital shaker.

31. Pour off the ethidium bromide stain and add 50 ml of ddH₂O. De-stain for 5 minutes with gentle mixing on a benchtop orbital shaker.

32. Pour off the de-stain ddH₂O and add a small amount of fresh ddH₂O to the dish. The fresh ddH₂O makes it easier to remove the gel from the dish. Don’t worry about adding too much.

At this point, the gel can be imaged on a gel-doc or UV scanner prior to the gel extraction step.

33. Prepare a 2.0 ml microcentrifuge tube containing 600 μl of 0.4 M ammonium acetate for each gel lane loaded.

34. Image the gel on a UV box. The ssDNA product of interest is 59 bases long. Excise the bands of interest and transfer to them to the microcentrifuge tubes prepared in step 33 (one gel slice/tube).

Some fluorophores may slightly alter the migration of the ssDNA relative to the DNA ladder. Three prominent bands are expected: one containing unlabeled strand and any un-nicked labeled DNA resulting from incomplete digestion; a faster migrating band containing the labeled probe DNA; the fastest migrating band containing the unincorporated primer sequences and the unlabeled fragment released by the nicking reaction.

Note: The final labeled ssDNA fragment of interest reported in Beliveau et al. (2012) is 53 bases long, while the fragment described here is 59 bases. The difference in length is the result of the inclusion of an additional 6-base nicking.
endonuclease recognition site that allows the isolation of either strand of the duplex produced by PCR (see Figure 1).

35. Incubate the gel slices overnight at 55°C in a heated vortexer at max speed (if available) or a shaking incubator at ≥300 rpm.

36. The next day, spin the tubes containing the gel slices for 5 minutes at max speed in a microcentrifuge.

37. Collect the supernatant from each tube, leaving the gel slice behind, and transfer to a fresh 2.0 ml microcentrifuge tube.

The gel slices can be incubated at 55°C with shaking in an additional 600 µl of 0.4 M ammonium acetate to recover any product not eluted by the first overnight incubation if desired.

38. Set up a DNA precipitation: add 13.5 µl 20 mg/ml glycogen and 1350 µl ice-cold 100% ethanol to each 2.0 ml tube. Vortex vigorously to mix.

39. Incubate the precipitations at ~80°C for 35 minutes or for >2 hours at ~20°C.

Precipitations can be left at ~20°C overnight or indefinitely; this step represents a convenient stopping point.

40. Spin the precipitation at max speed for 1 hour at 4°C in a refrigerated centrifuge.

The precipitation can also been spun at room temperature if a refrigerated centrifuge is not available.

41. Carefully aspirate off the 100% ethanol using a vacuum trap or micropipette.

A prominent pellet should be visible after the spin. Depending on the fluorophore used, the pellet may appear colored.

42. Add 1350 µl ice-cold 70% to each tube, taking care not to disturb the pellet.

43. Spin the precipitation at max speed for 30 minutes at 4°C in a refrigerated centrifuge.

44. Carefully aspirate off the 70% ethanol using a vacuum trap or micropipette.

45. Air dry the pellets by placing open tubes onto a 42°C heat block for 15 minutes.

46. Add 10 µl of molecular biology grade water to each pellet. Incubate for 60 minutes at 37°C to resuspend. Vortex each tube after the 37°C incubation, then spin briefly in a microcentrifuge.

Shaking at 1400 rpm in a heated vortexer will speed up the resuspension.

47. Quantify the amount of ssDNA FISH probe using a spectrophotometer.

The concentration of the ssDNA FISH probe can be inferred by measuring the absorbance of light at 260 nm (A_{260}). Remember to use the ssDNA absorbance unit of 37 µg/ml, not the 50 µg/ml absorbance unit of dsDNA.

It is also important to note that some fluorophores, particularly those in the blue to green portion of the visible spectrum, absorb light at 260 nm. Thus, to get an accurate DNA concentration, a Correction Factor (CF) may need to be used. CF is defined as \( \frac{A_{260 \text{ free dye}}}{A_{\text{max free dye}}} \), or the fluorophore’s absorbance as a free dye at 260 nm divided by the fluorophore’s absorbance as a free dye at its absorbance maximum (i.e. excitation wavelength). Therefore, the corrected DNA absorbance is expressed as \( A_{\text{DNA}} = A_{260} - (A_{\text{max}} \times \text{CF}) \), where \( A_{260} \) is the absorbance of the
solution at 260 nm and $A_{\text{max}}$ is the absorbance of the solution at the absorbance max of the fluorophore.

Probe preparations can be stored at 4°C for several weeks or −20°C indefinitely. Aliquot large preparations before freezing to minimize the number of times the probe is freeze-thawed.

**BASIC PROTOCOL 2 (optional)**

**Interphase FISH using Oligopaint probes**—Oligopaint probes label interphase nuclei extremely efficiently. This section details a fast FISH protocol optimized for labeling chromosomal DNA in fixed tissue culture cells. Alternate Protocol 2 describes a modified 3D-FISH (Lanzuolo et al., 2007; Cremer et al., 2008) version of this basic protocol in which the temperature is never raised above 78°C, and Alternate Protocol 3 presents a modified protocol for applying FISH with Oligopaint probes to metaphase chromosome preparations.

**Materials**

- Fixed interphase cells adhered to a glass microscope slide (see Support Protocol 2)
- 4X SSCT (see recipe)
- Formamide (store at 4°C, away from light)
- 2X hybridization cocktail (see recipe)
- 10 mg/ml RNase A
- Oligopaint probe (see Basic Protocol 1)
- 1.7 ml microcentrifuge tubes
- 22 × 22 mm #1.5 coverslips
- Rubber cement
- 2X SSCT (see recipe)
- 0.2X SSC (see recipe)
- Anti-fade mounting media with DAPI
- 22 × 30 mm #1.5 coverslips
- Nail polish
- 100 ml graduated cylinder
- Plastic paraffin film
- Glass ‘Coplin’ slide staining jars
- Adjustable temperature water bath (x2)
- Anodized aluminum heat block
- Forceps
- Benchtop vortexer
- Benchtop microcentrifuge
- Hybridization chamber (see step 15)
- Heated incubator
- Epifluorescent or confocal microscope

_Curr Protoc Mol Biol. Author manuscript; available in PMC 2015 January 06._
1. Prepare 100 ml of 2X SSCT + 50% formamide by adding 50 ml of 4X SSCT and 50 ml of formamide to a 100 ml graduated cylinder. Seal the top of the cylinder with plastic paraffin film and invert several times to mix.

2. Add 50 ml of 2X SSCT + 50% formamide per jar to two 'Coplin' glass slide staining jars. Place each Coplin jar in a separate temperature-adjustable water bath at room temperature.

   Caution: do not place a room temperature Coplin jar directly into a hot water bath, as this may weaken the Coplin jar or cause it to break instantly.

3. Heat one water bath to 60°C and the other to 92°C. Also place an anodized aluminum block into the bath to be set to 92°C and adjust the water level such that all but about 1 cm of the block is submerged. Allow the sample slides to warm from 4°C to room temperature while the water baths heat up.

4. Transfer the sample slides into the Coplin jar containing 2X SSCT + 50% formamide at 92°C using forceps and incubate for 2.5 minutes.

5. Transfer the slides into the Coplin jar containing 2X SSCT + 50% formamide at 60°C using forceps and incubate for 20 minutes.

6. While the slides are incubating at 60°C, prepare a hybridization master mix. For each sample, add 12.5 μl of 2X hybridization cocktail (Reagents and Solutions), 12.5 μl of formamide, and 1 μl of 10 mg/ml RNase. For each sample, aliquot 26 μl of hybridization master mix into a 1.7 μl microcentrifuge tube.

   The 2X hybridization cocktail is very viscous and can be difficult to pipette. Removing 1–2 cm from the end of a plastic pipette tip with scissors or a razor blade will make this step easier. If available, a positive-displacement pipette simplifies the pipetting.

7. Add the Oligopaint probe to each microcentrifuge tube containing hybridization master mix. 20 – 30 pmol of Oligopaint probe is typically sufficient to produce strong staining in fixed tissue culture cells; 10-fold more probe is recommended for tissue sections and whole mount tissues.

   The amount of probe needed to produce strong signal may need to be determined empirically for each type of sample. See the Critical Parameters section for further discussion of this issue.

   Try to keep the volume of probe added as small as possible - i.e. use concentrated stocks of probe. Dilute probes can be concentrated by lyophilization or salt-ethanol precipitation followed by resuspension in a reduced volume. Adding ≤~4 μl of probe per sample is recommended.

8. Mix the contents of each microcentrifuge tube by vortexing and spin down briefly in a benchtop microcentrifuge to collect the hybridization mix. Protect the hybridization mix from light by covering the tubes with foil or placing them in a drawer until the slides are ready.

9. After the 20 minute incubation at 60°C is finished, carefully remove the slides from the Coplin jar in the water bath using forceps. Partially dry the sample slides by tapping the thin edge of the slides against a paper towel, then place the slides into an empty Coplin jar.

   Take care that the slide surface containing the sample never comes into direct contact with the paper towel, as the paper towel can introduce unwelcome debris.
10. Place a 22 × 22 mm coverslip onto a dry paper towel and pipette the hybridization mix directly onto the coverslip.

The final hybridization mix will be less viscous than the 2X hybridization cocktail and can be pipetted using a standard micropipette with standard tips. Take care to avoid air bubbles when pipetting.

11. Gently invert the sample slide onto the coverslip such that the sample (typically visible as a cloudy circle) is placed in direct contact with the hybridization mix. Do not press the slide down against the coverslip as this will result in the ejection of hybridization mix onto the paper towel.

12. Turn the slide over so the coverslip is on the upward-facing side of the slide. Seal the coverslip onto the slide by adding a layer of rubber cement around the edges of the coverslip.

Do not worry about adding too much rubber cement as it is easily removed after the denaturation and hybridization steps. It is essential that the perimeter of the coverslip be completely covered by rubber cement, else the hybridization mix may leak or evaporate during the denaturation or hybridization step.

13. Allow the rubber cement to dry for 5 minutes.

The slides can be placed inside a box or in a drawer during this incubation to protect them from light.

14. Denature each slide by placing it coverslip-side up on the top of the submerged anodized aluminum block in the 92°C water bath for 2.5 minutes.

Take care not to let the slides sit on the block for longer than 2.5 minutes as the sample may begin to be destroyed by extended high-temperature treatment. If the size of the water bath and the number of aluminum blocks available permits it, several samples can be denatured in parallel.

15. Transfer the denatured slide to a humidified hybridization chamber and allow the hybridization reaction to occur overnight (>14 hours) at 42°C in a heated incubator.

A humidified hybridization chamber can be assembled from a plastic chamber with a lid (e.g. a plastic Tupperware container or an empty pipette tips box), damp paper towels, and a plastic rest for the slides (to keep them from sitting directly on the damp paper towels) such as a reagent reservoir or few pieces of a cut serological pipette. The chamber does not need to be airtight and preferably will have the lid only loosely attached to allow for air exchange when placed in the heated incubator.

16. The next day, add 50 ml of 2X SSCT to a Coplin jar. Place the Coplin jar in a water bath and heat to 60°C.

If your anti-fade mounting media is stored at −20°C, remove it from the freezer at this point and allow it to warm to room temperature.

17. Carefully remove the rubber cement and coverslips from the sample slides and transfer to the Coplin jar containing 2X SSCT in the 60°C water bath. Incubate for 15 minutes at 60°C.

The coverslip can often by removed by gently sliding it along the surface of the slide towards the edge using a gloved finger. If the coverslip does not move readily, remove the rubber cement with fine forceps while applying gentle pressure to the center of the coverslip. In both methods, it is essential to avoid violently prying the coverslip upwards (this often occurs when just one edge is freed from rubber
cement but its parallel edge is still covered), as this will result in the stretching/shearing of chromosomal DNA.

18. Transfer the slides to a Coplin jar containing 50 ml of 2X SSCT and incubate for 10 minutes at room temperature.
   The Coplin jar can be covered with foil or placed in a drawer to protect the samples from light.

19. Transfer the slides to a Coplin jar containing 50 ml of 0.2X SSC and incubate for 10 minutes at room temperature.
   The Coplin jar can be covered with foil or placed in a drawer to protect the samples from light.

20. Remove the slides from the 0.2X SSC Coplin jar and partially dry the slides by tapping the thin edge of the slides against a paper towel, then placing the slides into an empty Coplin jar.

21. Place a 30 × 22 mm coverslip onto a dry paper towel and pipette 12.5 μl of anti-fade mounting media containing DAPI onto the center of the coverslip. Take care to avoid air bubbles.

22. Gently invert the sample slide onto the coverslip such that the sample (typically visible as a cloudy circle) is placed in direct contact with the mounting media. Press down gently but firmly to eject excess mounting media.
   The coverslip should be immobile to the touch, being held in place by a thin monolayer of mounting media. If the coverslip is still loose, move the inverted slide + coverslip to another spot on the paper towel and repeat. Take care not to press too firmly as the slide may break.

23. Turn the slide over so the coverslip is on the upward-facing side of the slide. Seal the coverslip onto the slide by adding a layer of nail polish around the edges of the coverslip.

24. Allow at least 30 minutes for the nail polish to dry before imaging the slides. The FISH can be visualized using an epifluorescent or confocal microscope.

ALTERNATE PROTOCOL 2 (optional)

Fast 3D-FISH using Oligopaint probes—This section presents a relatively quick 3D-FISH (Lanzuolo et al., 2007; Cremer et al., 2008) protocol that avoids using high-temperature treatments in order to better preserve the nuclear morphology of samples. This protocol is similar to Basic Protocol 2, but contains additional pre-hybridization steps to improve sample permeability.

Additional Materials

1X PBS (see recipe)
1X PBST (see recipe)
1X PBS + 0.5% (vol/vol) Triton-X100
0.1 N HCl

1. Heat one water bath to 60°C and the other to 78°C. Place a Coplin jar containing 50 ml of 2X SSCT + 50% formamide (Basic Protocol 2 steps 1–2) into the 60°C bath and place an anodized aluminum block into the 78°C bath and adjust the water
level such that all but about 1 cm of the block is submerged. Allow the sample slides to warm from 4°C to room temperature while the water baths heat up.

Note: unless otherwise indicated, perform all of the following incubations at room temperature.

2. Transfer the sample slides using forceps to a Coplin jar containing 50 ml of 1X PBS and incubate for 1 minute.

3. Transfer the sample slides to a Coplin jar containing 50 ml of 1X PBST and incubate for 1 minute.

4. Transfer the sample slides to a Coplin jar containing 50 ml of 1X PBS + 0.5% (vol/vol) Triton-X100 (250 μl in 50 ml) and incubate for 10 minutes.

5. Transfer the sample slides to a Coplin jar containing 50 ml of 1X PBST and incubate for 2 minutes.

6. Transfer the sample slides to a Coplin jar containing 50 ml of 0.1 N HCl (625 μl 8N HCl in 50 ml ddH$_2$O) and incubate for 5 minutes.

7. Transfer the sample slides to a Coplin jar containing 50 ml of 2X SSCT and incubate for 1 minute.

8. Transfer the sample slides to a Coplin jar containing 50 ml of 2X SSCT and incubate for 2 minutes.

9. Transfer the sample slides to a Coplin jar containing 50 ml of 2X SSCT and incubate for 2 minutes.

10. Transfer the sample slides to a Coplin jar containing 50 ml of 2X SSCT + 50% formamide and incubate for 5 minutes.

11. Transfer the sample slides to the Coplin jar containing 50 ml of 2X SSCT + 50% formamide in the 60°C water bath and incubate for 20 minutes.

12. Follow Steps 6–24 of Basic Protocol 2 exactly as written, except denaturing at 78°C for 2.5 minutes in Step 14.

ALTERNATE PROTOCOL 3

Metaphase FISH with Oligopaint probes—Metaphase FISH with Oligopaint probes calls for a slightly modified protocol that uses different denaturation conditions and includes a progressive ethanol dehydration prior to probe addition.

Additional Materials

Sample slide containing spread mitotic chromosomes
70% (vol/vol) ethanol in ddH$_2$O
90% (vol/vol) ethanol in ddH$_2$O
100% (vol/vol) ethanol

1. Place a Coplin jar containing 50 ml of 2X SSCT + 70% (vol/vol) formamide in a water bath and warm to 70°C.

2. While the water bath is heating up, allow sample slides to warm to room temperature.
3. Denature the samples by incubating them in the Coplin jar containing 2X SSCT +
70% formamide at 70°C for 1.5 minutes.
4. Transfer the samples to a Coplin jar containing ice cold 70% ethanol and incubate
at room temperature for 5 minutes.
5. Transfer the samples to a Coplin jar containing ice cold 90% ethanol and incubate
at room temperature for 5 minutes.
6. Transfer the samples to a Coplin jar containing ice cold 100% ethanol and incubate
at room temperature for 5 minutes.
7. Air dry the samples face up on a paper towel at room temperature (1-2 minutes).
The sample slides can be placed inside a box or drawer to protect them from
floating dust.
8. Prepare hybridization mix as directed in Steps 6-8 of Basic Protocol 2.
9. Add the hybridization mix to the sample slides and seal with rubber cement as
directed in Steps 10-12 of Basic Protocol 2.
10. Transfer the sample slides to a humidified hybridization chamber and allow the
hybridization reaction to occur overnight (>14 hours) at 37°C in a heated incubator.

SUPPORT PROTOCOL 2 (optional)

Preparing tissue culture cells for interphase FISH—Tissue culture cells provide a
convenient substrate for interphase FISH. This section describes a standard protocol for
creating FISH sample slides from a suspension of adherent or semi-adherent cells.

Additional Materials

- 0.01% (vol/vol) poly-L-Lysine solution in ddH2O
- 1×10^5 - 1×10^6 cells/ml cell suspension
- 1X PBS + 4% (vol/vol) paraformaldehyde (see recipe)
- 25 × 75 × 1 mm glass microscope slides
- Lint-free paper towels (e.g. Kimwipes)
- Plastic Coplin staining jar
- Cell culture incubator

Note: All steps are performed at room temperature unless otherwise indicated
1. Use a lint-free paper towel soaked in 100% ethanol to clean each slide
   This step is recommended for all slides, even those advertised as “pre-cleaned” by
   their manufacturer, and is intended to remove microscopic dust and debris from the
   slides.
2. Allow the slides to air dry.
3. Prepare 50 ml of 0.01% (vol/vol) poly-L-Lysine solution in ddH2O (5 ml 0.1% stock + 45 ml ddH2O) and add to a plastic Coplin jar.
   A glass Coplin jar can be used here, but its insides will be coated by the poly-L-
   lysine solution as well.
4. Allow the slides to air dry.  
   This step can be sped up using the airflow of a tissue culture hood.
5. Prepare a suspension of $1 \times 10^5$ - $1 \times 10^6$ cells/ml in complete growth media.  
   This density works for most cell lines but can be adjusted if necessary.
6. Use a micropipette to place 100 µl of cell suspension roughly in the center of a 
   poly-L-lysine-treated slide.
7. Allow the cells to adhere for 1.5 - 3 hours in a cell culture incubator.  
   Large Petri dishes and metal trays covered with aluminum foil are convenient 
   vessels to hold the slides. If working with mammalian cells, be sure to allow for gas 
   exchange between the slides and the atmosphere of the incubator.
8. Transfer the slides to a Coplin jar containing 50 ml of 1X PBS and incubate for 1 
   minute.
9. Transfer the slides to a Coplin jar containing 40 ml of 1X PBS + 4% (vol/vol) 
   paraformaldehyde and incubate for 10 minutes.
10. Transfer the slides to a Coplin jar containing 50 ml of 1X PBS and incubate for 1 
    minute.
11. Transfer the slides to a Coplin jar containing 50 ml of 2X SSCT and incubate for 5 
    minutes.
12. Transfer the slides to a Coplin jar containing 50 ml of 2X SSCT + 50% formamide 
    (25 ml 4X SSCT + 25 ml formamide) and incubate for 5 minutes.
13. Transfer the slides to a Coplin jar containing 50 ml of 2X SSCT + 50% formamide 
    and transfer to 4°C for storage.  
    Slides should be used within 2 weeks of creation. Screw-top Coplin jars are 
    convenient for storing slides; paraffin film can be used to seal the top of standard 
    Coplin jars if screw-top jars are not available.

REAGENTS AND SOLUTIONS

Note ddH$_2$O refers to deionized, distilled water.

2X hybridization cocktail

For 10 ml: add 4 ml 50% (wt/vol) dextran sulfate solution (in ddH$_2$O), 2 ml 20X SSC, and 4 
ml ddH$_2$O. Store at room temperature.

1X PBS—Dilute 10X PBS (see recipe) 1:10 in ddH$_2$O. Store at room temperature.

10X PBS—Prepare a 1 L 10X PBS stock solution: dissolve 80 g NaCl, 2 g KCl, 14.4 g 
Na$_2$HPO$_4$, and 2.4 g KH$_2$PO$_4$ in 800 ml ddH$_2$O. Adjust pH to 7.4 using HCl, and then add 
ddH$_2$O to 1 L. Store at room temperature.

1X PBS + 4% (vol/vol) paraformaldehyde—For 40 ml: combine 4 ml 10X PBS, 26 ml 
ddH$_2$O, and 10 ml of 16% (wt/vol) paraformaldehyde solution (made with ddH$_2$O). Make 
fresh and store at 4°C for up to one week.

1X PBST—For 1 L: combine 100 ml 10X PBS, 899 ml ddH$_2$O, and 1 ml Tween-20. Store 
at room temperature.

_Curr Protoc Mol Biol_. Author manuscript; available in PMC 2015 January 06.
0.2X SSC—For 1 L: combine 20 ml 20X SSC and 980 ml ddH₂O. Store at room temperature.

20X SSC—For 1 L: dissolve 175.3 g of NaCl and 88.2 g of sodium citrate in 800 ml of ddH₂O. Adjust the pH to 7.0 with 1M HCl, and then add ddH₂O to 1 L. Store at room temperature.

2X SSCT—For 1 L: combine 200 ml 20X SSC, 799 ml ddH₂O, and 1 ml Tween-20. Store at room temperature.

4X SSCT—For 1 L: combine 400 ml 20X SSC, 598 ml ddH₂O, and 2 ml Tween-20. Store at room temperature.

8X SSCT—For 1 L: combine 800 ml 20X SSC, 196 ml ddH₂O, and 4 ml Tween-20. Store at room temperature.

1X TBE—Prepare 1L of a 5X stock solution: dissolve 54 g of Tris base, 27.5 g of boric acid, and 20 ml 0.5 M Ethylenediaminetetraacetic acid (EDTA) pH 8.0 in ddH₂O. Dilute the 5X stock 1:5 in ddH₂O to make 1X. Store at room temperature.

COMMENTARY

Background Information

FISH is an adaptation of a nucleic acid in situ hybridization method that was developed by Pardue and Gall (Pardue and Gall, 1969) and used radiolabeled probes (Bauman et al., 1980; Pinkel et al., 1986; reviewed in: Levsky and Singer, 2003; Volpi and Bridger, 2008; Itzkovitz and van Oudenaarden, 2011). It is a powerful tool for studying chromosome structure and position as well as gene expression, as it can be used to interrogate both DNA and RNA molecules. It is informative at the single-cell level and can be combined with other single-cell imaging methods, such as the immufluorescent detection of proteins.

FISH probes have typically been derived from flow-sorted chromosomes or genomic inserts subcloned into vectors, such as plasmids, cosmids, and BACs. These chromosomes and inserts can be labeled via the incorporation of fluorophore- or hapten-conjugated dNTPs in enzymatic nick-translation or PCR reactions, producing ~100 bp dsDNA (Lichter et al., 1988). Because the starting material can contain sequences that are highly repeated in the genome, the resulting probes often need to be used in the presence of unlabeled repetitive DNA so as to minimize background signal due to off-target hybridization (Landegent et al., 1987).

Synthetic nucleic acid oligos, including DNA, RNA, peptide nucleic acid (PNA), and locked nucleic acid (LNA) oligos, have also been used as FISH probes (Larsson et al., 1988; Landsdorp et al., 1996; O'Keefe, Warburton, and Matera, 1996; Silahtaroglu, Tommerup, and Vissing, 2003). Such oligo probes have typically been used to detect RNA molecules (Femino et al., 1998; Player et al., 2001; Raj et al., 2008) or to visualize DNA targets that facilitate signal detection because they are internally repetitive (Dernburg et al., 1996). In these situations, the number of oligo species in a probe has hovered in the range of ~1 to 50.

Excitingly, recent technological advances permitting the parallel synthesis of hundreds to hundreds of thousands of oligo species have enabled several new methods for generating oligo-based FISH probes. One family of these methods uses synthetic oligos containing genomic sequence as PCR primers to amplify fragments of genomic sequence in parallel reactions (Martinez et al., 2006; Navin et al., 2006; Lamb et al., 2007; Bienko et al., 2013).
Other methods encode the entirety of the genomic region to be targeted in a library of oligos and, therefore, permit the pool of oligos to be used directly as a FISH probe, if fluorescent label is added during synthesis of the library (Boyle et al., 2011), or, if the genomic sequences are flanked by primers, enable the pool to become a renewable resource that can be amplified in the presence of fluorescent label to generate dsDNA (Yamada et al., 2011; Beliveau et al., 2012) or ssDNA (Beliveau et al., 2012) probes. Note that the use of primers that have been synthesized to contain precisely positioned fluorophores (e.g., 5' labeled primers) will produce probes with uniform specific activities of fluorescence (Beliveau et al., 2012). The strategy of Beliveau et al., 2012, which produces Oligopaints, is compatible with many standard molecular biological protocols and can be used to generate a variety of probe structures (see Beliveau et al., 2012 Fig. S12). In this unit, we detail the version that produces probes which are short (e.g. 59 bases) and single-stranded, features that likely enhance the diffusion of probes into samples as well as increase the efficiency of hybridization. The strand-specific nature of Oligopaints will also aid in the design of experiments that target RNA molecules or must distinguish one strand of DNA from the other.

Critical Parameters

**Probe design**—In general, targeting genomic regions with >500-1000 probes, where each oligo carries one fluorophore, is sufficient to produce a strong signal. The number of oligo species required to produce a robust FISH signal may, however, vary depending on the genomic target and the type of sample (e.g., cell culture or tissue sample) being interrogated. If the region of interest (e.g. a gene body) is too small to accommodate the desired number of probe oligos, researchers may wish to extend the target into flanking sequences. In cases where extending the target is not desirable, users may wish to consider methods for amplifying signals, such as placing a hapten (e.g. biotin or digoxigenin) on the primer and then visualizing the FISH using antibodies that target the hapten. The design of probes for RNA FISH may also vary greatly depending on the target; interested readers are encouraged to read the discussion of this topic in Itzkovitz and van Oudenaarden (2011).

**Reagent Quality**—The generation of Oligopaint probes using the method outlined in this unit relies heavily on two enzymatic steps – PCR and digestion with a nicking endonuclease. It is important to monitor the quality of the reagents used in these steps, as one bad tube of enzyme, buffer, or stock of molecular biology grade water can destroy an entire preparation. Some general guidelines for handling these reagents are listed below:

- Take care to store enzymes at the temperature recommended by their manufacturer (typically −20°C) and always store the enzymes in a cooler/caddy when they are not in a freezer.
- Avoid repeated freeze/thaw cycles. Using a laboratory grade marker, place a dot on the top of the tube each time it has been thawed to keep track of the number of times a tube has been freeze/thawed. Aliquot larger stocks if necessary.
- Only use molecular biology grade water in enzymatic reactions; make 15 ml aliquots by decanting from the stock bottle into conical tubes in a tissue culture or PCR hood to avoid having debris fall into the stock bottle. Date each aliquot and use within 2 weeks of opening for the first time.

**Physical Mixing of Reactions**—Our protocol for generating Oligopaints requires that commonly used molecular biological techniques be performed at scales that may be significantly larger than is customary for many users. As such, we emphasize the importance of vortexing large master mixes (e.g. for PCR, DNA precipitation, nicking endonuclease digestion) enough to ensure homogeneity. In our hands, master mixes containing
thermostable DNA polymerase and the Nb.BspDI and Nb.BsmI nicking enzymes are robust to repeated (e.g. 3–5X) 10–15 second full-speed vortex pulses.

**Running Gels under Denaturing Conditions**—It is critical to run the denaturing gel at a temperature that will maintain separation of the desired labeled ssDNA probe molecules from their complement (~55–60°C, for the conditions described in the protocol above). Note that a given probe molecule will have partial complementarity to many DNA molecules after denaturation by virtue of the retained primer sequence; thus, partial reannealing may occur even between molecules that have heterologous genomic sequence inserts. If the gel is not sufficiently hot, slow-migrating smears may be present above both the band of the desired ssDNA product and the band of the full length uncut strands. Note that a labeled ssDNA fragment can still be successfully purified if it has migrated far enough from the smeared DNA to allow for its unambiguous isolation.

**Probe concentration**—While Oligopaint probes tend to work robustly at oligo concentrations of 800 nM or higher (20+ pmol per 25 μl hybridization reaction) for tissue culture cells, users are encouraged to determine the optimal concentration for their specific needs via a concentration curve. For example, increasing the concentration of probe in the hybridization reaction may be necessary should an 800nM concentration of Oligopaint probes be insufficient to produce a strong signal. As for tissue samples, increasing the concentration of probe by up to 10-fold or more may be necessary, with the optimal concentration likely to vary depending on the sample type and target.

**Conditions for hybridization and washes**—The hybridization and wash conditions can be altered to optimize the performance of the FISH probes. Both hybridization temperature and wash conditions can impact the signal:noise of the FISH (see Beliveau et al., 2012 Fig. S6, S8). For example, higher hybridization temperatures or higher temperatures in the first wash may reduce the background signal. The length of the hybridization step can also be important; >14 hours is recommended, and >24 hours can be tried for probes that do not produce strong signals after 14 hours. The signal:noise of the FISH can also be adjusted by increasing the number of high temperature wash steps (e.g. 2x 15’ 60°C 2X SSCT), increasing the temperature of the first wash (e.g. 70°C SSCT), or adding formamide to the first wash step (e.g. instead 15’ 37°C 2X SSCT + 50% (vol/vol) formamide instead of 15’ 60°C 2X SSCT).

**Sample Permeability for FISH**—Some samples may require permeabilization steps beyond the pre-hybridization treatments included in Basic Protocol 2. Common options include incubation in detergents such as tween-20 or triton X-100, treatment with proteases such as pepsin and proteinase K, treatment with HCl, flash-freeses in liquid nitrogen, and fixation in 3:1 (vol/vol) methanol:acetic acid instead of paraformaldehyde (Lanzuolo et al., 2007; Cremer et al., 2008). For samples that do not work readily in Basic Protocol 2, users can try performing Steps 1–11 of Alternate Protocol 2 and then following Steps 6–24 of Basic Protocol 2.

**Anticipated Results**

A successful Oligopaint probe preparation typically converts 20–30% of the labeled primer into purified ssDNA FISH probe. This yield should be reasonably consistent across different preparations of the same Oligopaint probe as well as preparations of distinct Oligopaint probes; and, in our hands, appears to be largely insensitive to the specific fluorophores we have used (Alexa405, 6-FAM, Alexa488, atto550, Cy3, TYE563, atto565, Cy5, atto647N, Alexa647, TYE665, Cy5.5).
FISH performed according to Basic Protocol 2 with a probe complexity of 1000 oligos or more and a total oligo concentration of 800 nM should produce a signal in the nucleus of >90% and often close to 100% of tissue culture cells. FISH efficiency may vary depending on the genomic target, the complexity and concentration of the probe, and the nature of the sample.

**Time Considerations**

Oligopaint probe preparations can be completed in as few as three days or spread over the course of a week.

Preparing sample slides according the Alternate Protocol 2 takes approximately 3–4 hours.

FISH with Oligopaint probes takes approximately 4 hours spread over two days: two hours on the first day to perform the pre-hybridization and denaturation steps and set up the overnight hybridization, and two hours on the following day to perform the wash steps, mount the slides, and perform microscopy.

**Acknowledgments**

The writing of this unit was aided by critical discussions with several members of the C.-ting Wu laboratory: Frederic Bantignies, E. Joyce, T.N. Senaratne, M. Hannan, C. Fonseka, R. McColle, and S. Nguyen. We also benefited from valuable conservations with A. Boettiger, J. Moffit, T. Schmidt, C. Kim-Kiselak, S. Elledge, R. Kingston, and D. Moazed. The development of Oligopaints was supported by a National Institutes of Health (NIH)/National Institute of General Medical Sciences Grant (1R01GM085169) and Pioneer Award, a Broad Institute Scientific Planning and Allocation of Resources Committee (SPARC) Award, and a Cox Program Award from Harvard Medical School.

**LITERATURE CITED**


Carr Protoc Mol Biol. Author manuscript; available in PMC 2015 January 06.
Figure 1. Library design using the Oligopaints strategy

Each synthetic ssDNA library may contain hundreds to hundreds of thousands or more unique oligo species and can contain multiple distinct Oligopaint probes through the use of multiple primer pairs. This figure diagrams the structure of two oligos from different probe sets in a hypothetical library: an oligo belonging to a probe that can be amplified using primer pair 1 (top left) and an oligo belonging to a probe that can be amplified using primer pair 2 (bottom left). Both oligos contain, in 5’ to 3’ order, the sequence of the forward primer, genomic sequence, and then the reverse complement (RC) of the reverse primer. The sites for two nicking endonucleases, Nb.BsmI and Nb.BsrDI, are placed between the primer sequences and the genomic sequences. The use of two nicking endonuclease sites in the library molecules allows for the production of strand specific probes: amplification with a labeled F primer and digestion with Nb.BsrDI will yield probe targeting the reverse complement of the genomic sequence encoded in the ssDNA molecules of the library, whereas amplification with a labeled R primer and digestion with Nb.BsmI will yield probe targeting the genomic sequence encoded in the ssDNA molecules of the library. The use of two nicking endonuclease sites, instead of one, is an update of the strategy presented in Figure 1 of Beliveau et al. (2012).
Figure 2. Examples of FISH with Oligopaint probes

(A) FISH using an Oligopaint probe composed of 20,020 oligos targeting a 2.5 Mb region on human chromosome X (Xq13.1) in diploid (XY) MRC-5 cells. Hybridization was carried out according to Basic Protocol 2 in a 25 µl volume using 35 pmol of probe (Cy3, red) produced using Basic Protocol 1. DNA is stained with DAPI (blue). (B) Two-color FISH on a field of tetraploid Drosophila Kc167 cells using two Oligopaint probes targeting the right arm of chromosome 2: one composed of 25,000 oligos targeting a 2.7 Mb region (50D1-53D7), and another composed of 50,000 oligos - half targeting a 3 Mb region (41E3-44C4) and the other half targeting a 2.6 Mb region (58D2-60D14). The hybridization was carried out according to Basic Protocol 2 in a 25 µl volume using 5 pmol of the first probe (Cy3, red) and 2.5 pmol of the second probe (Cy5, green), both produced using Basic Protocol 1. (C) Three-color FISH performed on Drosophila salivary polytene chromosomes with the two Oligopaint probes used in (B) plus an additional Oligopaint probe that also targets the right arm of chromosome 2 and is composed of 105,000 oligos - half targeting a 5.6 Mb region (44C4-50C9) and the other half targeting a 5.5 Mb region (53C9-58B6). The hybridization was carried out according to Beliveau et al. (2012) in a 25 µl volume using 20 pmol of each of the three probes labeled with TYE563 (red), TYE665 (blue), and 6-FAM (green), respectively, all produced using Basic Protocol 1. DNA is stained with DAPI (gray). Images represent confocal maximum Z projections (A, B) or a single confocal XY plane (C). Scale bars: 10 µM.