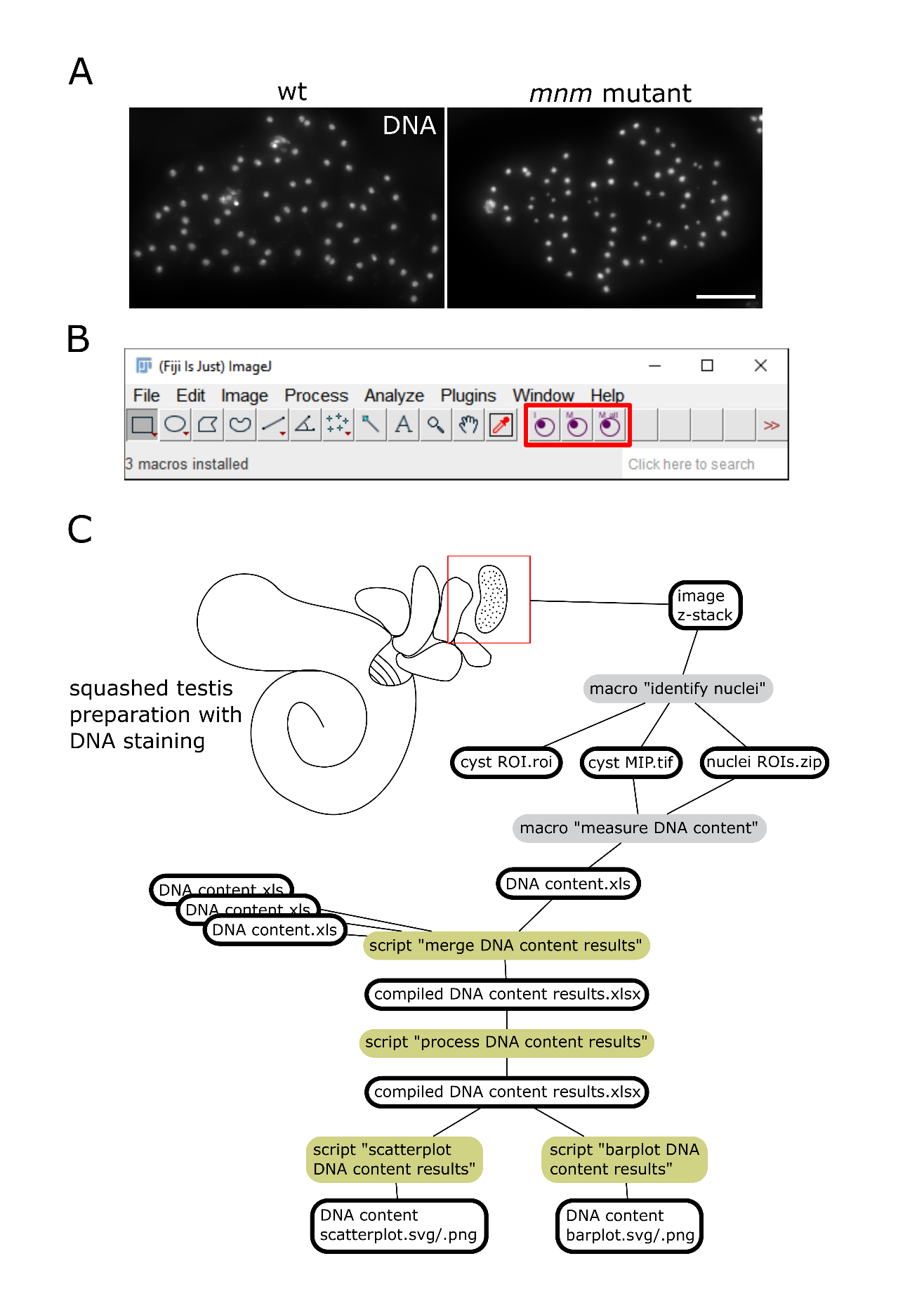
Complete extract from the PhD thesis (JW 2020)

****

**Figure 3.1: Pipeline for analysis of meiotic chromosome missegregation based on DNA content variation in early spermatid nuclei.**

(A) Maximum intensity projections of microscopic image stacks containing early postmeiotic cysts from testis squash preparations labeled with the DNA stain Hoechst 33258. Testes were isolated from either control (*w*1118) or *mnm* mutants (*w*\*;; *mnm*Z3-5578/ *mnm*Z3-3298). In the control, spermatid nuclei have similar sizes and signal intensities. In the mutants, spermatid nuclei vary considerably in size and signal intensity. Scale bar = 20 µm.

(B) Image of the main window of FIJI with the DNA content analysis toolset loaded. The individual macros of the toolset can be run by clicking on the icons highlighted with the red rectangle. The icon with an “I” runs the macro “identify nuclei”, the icon with an “M” runs the macro “measure DNA content” and the icon with the “M all” runs the macro “measure DNA content in subfolders”.

(C) Schematic representation of the DNA content analysis pipeline. The drawing represents a testis from an adult fly with cysts spilling out from a cut site. Files created throughout the pipelines are represented by their extension name and a black outline. ImageJ macros and Python scripts have background boxes that are grey and yellow-green, respectively. Lines down from a macro or a script lead to their output files. Lines down from a file to a macro or a script indicate their required input files.

## **Method 1:** A novel quantitative chromosome segregation assay based on microscopic DNA quantification in postmeiotic nuclei

### Overview

In contrast to the other assays, the analysis of DNA content in postmeiotic cysts has a drawback that can be overcome relatively easily. To make the assay quantitative one must convert the information contained in the images into numbers. Quantifying DNA can be done readily after staining with Hoechst 33258. This dye is highly specific for DNA, and signal intensities correlate well with the amount of DNA. After measuring the integrated Hoechst 33258 signal intensity in each spermatid of a cyst, the standard deviation describing the variation of the DNA content among the spermatid nuclei within the cyst provides an estimate of chromosome segregation error. In wild type cysts, the standard deviation should be near 0. Practically, deviations from 0 will arise only from the difference between male and female gametes and from technical variability. In contrast, with random segregation of chromosomes in meiosis I, large differences in DNA content between nuclei arise, sharply increasing the standard deviation. The first quantification protocol devised by Michael Sun and myself (Sun et al., 2019) contained several steps performed manually, including much of the result processing. Application of this original protocol was therefore difficult without in-depth knowledge of all the successive steps. Here, I describe a revised version, where all except very few steps are automatized, making its realization very simple (see protocol at the end of this part). The first part of the assay consists in making standard testis squash preparations for microscopy followed by staining with Hoechst 33258. Thereafter, image stacks from postmeiotic cysts are acquired with conventional wide-field fluorescence microscopy. These images are then analyzed using ImageJ macros coded specifically for this assay. Finally, the resulting data is processed and converted into plots using python scripts coded specifically for this assay. I have structured the code for this assay with two goals in mind: First, the assay can be performed without any knowledge of programming. Second, the code is optimized for human readability to allow people to understand it and change it according to their needs. The code is available as supporting information S3.1, S3.4, S3.5, S3.6, S3.7 and on github (https://github.com/BioJoe/AHC-assay-by-DNA-content).

### Requirements to run the DNA content analyses pipeline

The DNA content analysis pipeline consists of three ImageJ macros and four python scripts (Fig 3.1C). To use the ImageJ macros, a functional installation of FIJI (Schindelin et al., 2012) is required. The ImageJ macros are packed in the form of an ImageJ toolset. Macros in a toolset can be conveniently started from the toolbar in ImageJ (Fig 3.1B). To install a toolset in FIJI, a .txt file containing the code must be placed in “C:\Users\username\Fiji.app\macros\toolsets”*.* To use the python scripts, python 3.7 or later is required. Furthermore, four python packages need to be installed: “NumPy” and “pandas” are required for scientific computing, “Matplotlib” and “Seaborn” are required for data visualization.

### Slide preparation

Testis squash preparations should be prepared essentially as described previously (White-Cooper 2004; protocol 3.3.2). In brief, 5-10 pairs of testes are dissected from 1-3 days old males. On a poly-L-lysine-treated slide, the testes are cut open to release their content. Early postmeiotic cysts with round nuclei at the onion stage are generally found at about 1/3 of the length of the testis tube. Therefore, I recommend focusing on releasing cysts from this region. The testes are then gently squashed under a coverslip to spread the released cysts. After transient quick freezing in liquid nitrogen, testes are fixed using formaldehyde and stained for DNA using Hoechst 33258. The DNA content analyses can be performed in parallel with FISH, in that case the FISH staining should be performed as described in (Sun et al., 2019).

### Imaging

The image analyses pipeline has been set up using 1388x1040 pixels images from a Zeiss Cell Observer HS microscope with a 40×/0.75 oil immersion objective. The z-step size of the images was 280 nm and the number of slices was chosen in a way that all the nuclei from the cyst to be analyzed are fully included (usually 30-50 slices). Cysts of early post meiotic cysts with round nuclei at the onion stage are optimal to measure the DNA content. The hybridization process of FISH staining often results in fuzzy and more spread out DNA staining. The DNA content analyses is robust and works even with DNA staining of relatively low quality. However, if the difference in signal intensity between the nuclei and the background is too small, the recognition of nuclei fails. If more than 10 percent of the nuclei are not recognized the image should be discarded.

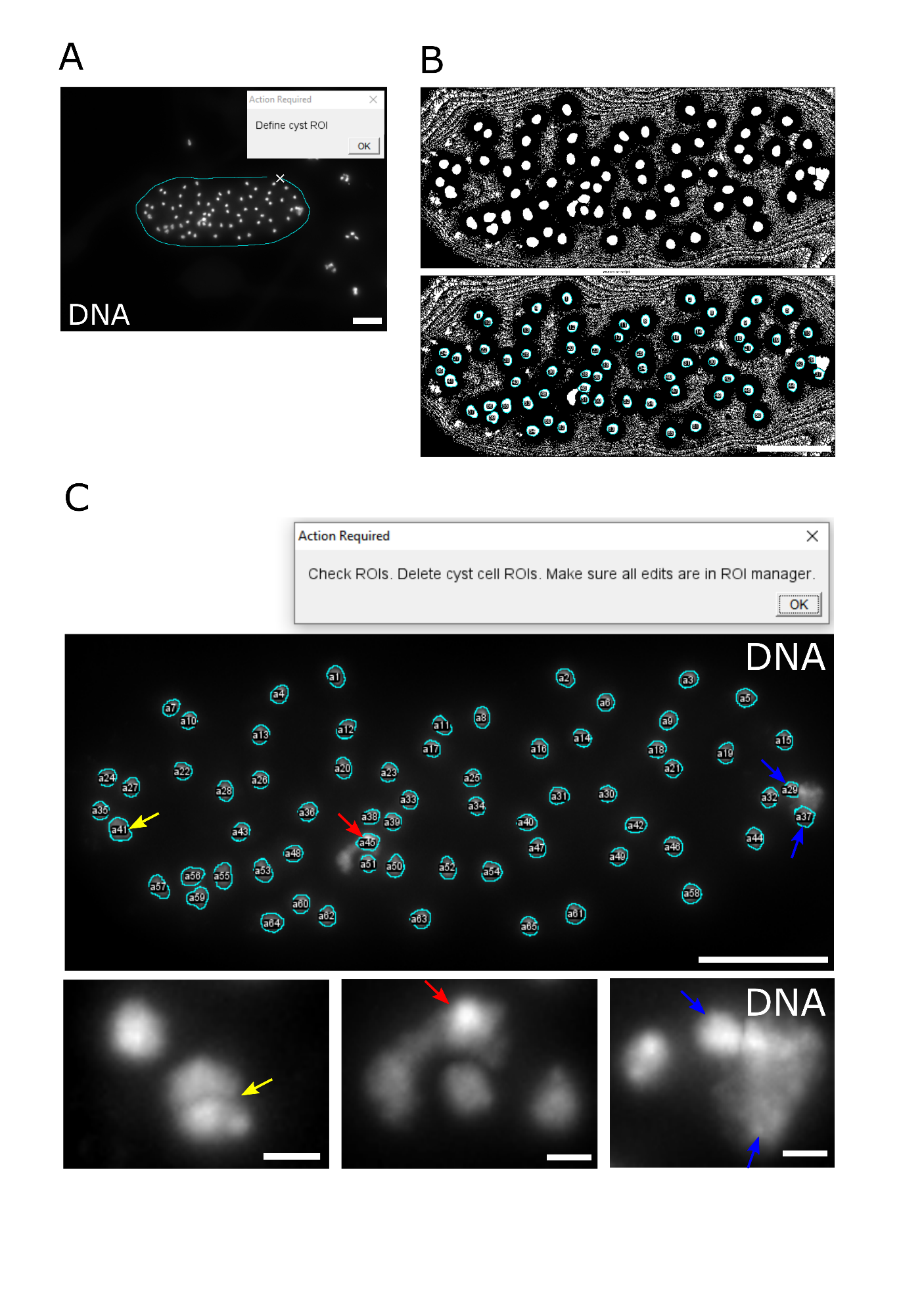
### Image analysis step 1: Identifying nuclei

The first step of the pipeline is the identification of postmeiotic nuclei as regions of interest (ROIs). The macro “identify nuclei” provides a user-friendly approach to a standard nuclei detection procedure with a few additional features tailored to this DNA content analyses pipeline. When running the macro, the user is first asked to provide the location of an image stack of a postmeiotic cyst as “.czi”. After loading the image stack a maximum intensity projection (MIP) is created and presented to the user, which is prompted to define the region of the cyst in which nuclei will be identified (cyst ROI) using the ImageJ freehand tool (Fig 3.2A). The defined cyst ROI is saved in the same folder and with the name of the original image stack but with the added name extension " cyst ROI.roi". All files saved throughout the analyses follow this convention. The MIP is then cropped to remove parts that are not in the cyst ROI and the dynamic range is reduced to the dynamic range of the pixels inside the cyst ROI. The MIP image is saved with the extension “ cyst MIP.tif”. All channels beside the DNA channel are removed. A binary image is created using a local mean thresholding algorithm. A watershed algorithm is then used to separate particles that touch each other (Fig 3.2B). Finally, on the binary image, nuclei are identified as ROIs using the "Analyze Particles” function (Fig 3.2B). Nuclei ROIs are renamed from a1 to an (n is the number of nuclei detected). The user is then prompted to review the nuclei ROIs and make changes if necessary (Fig 3.2C).

Manually defined ROIs are named u1-u*k* (*k* is the number of user-defined ROIs). The nuclei ROI set is saved with the extension " nuclei ROIs.zip".

The “Identify nuclei” macro has two steps of user intervention. The first requires the user to define a region of interest for the cyst. Ideally, the image contains at least one intact cyst, which is clearly separate from the surrounding tissue parts. In such images, it is straightforward to define the cyst ROI. However, even in good preparations such cysts are rare. Often the cysts are broken, with spermatids spilling out. I generally consider cysts worth analyzing if more than 50% of the spermatids can be analyzed.

The second step requiring user intervention is the correction of the nuclei ROIs. The automated nuclei detection does not differentiate between spermatid nuclei and cyst cell nuclei. Therefore, the user must manually remove the nuclei ROIs on cyst cells (red and lower blue arrow in Fig. 3.2C). Furthermore, if nuclei are positioned above each other in the z-dimension their signal will overlap in the MIP (yellow arrow in Fig. 3.2C). These overlapping nuclei are often recognized as a single ROI. Although not frequent enough to change the overall results, these ROIs should still be removed to reduce technical variability. Manually drawn ROIs can be added at this step. However, to prevent inducing user-bias, the user-defined ROIs are not taken into consideration when measuring DNA content. The user-defined ROIs are only used to identify FISH signals (see below for FISH analyses pipeline). After running this macro, three new files will have been added to the folder of the original image stack: the cyst ROI, the MIP and the nuclei ROIs (Fig 3.1).

**Figure 3.2: Automated identification of postmeiotic nuclei.**

(A) The first user-intervention when running the “identify nuclei” macro. The dialog box tells the user to outline a cyst ROI in the microscopic image, which displays a maximum intensity projection (MIP) of an early postmeiotic cyst after DNA staining. The cyan line is from the ImageJ freedraw tool that is being used to draw around the cyst. Scale bar = 20 µm.

(B) Binary images of the MIP shown in (A) are displayed. The image in the top panel was obtained after performing a local thresholding and watershed to identify nuclei. In the image of the bottom panel, the nuclei ROIs after identification by the “analyze particle function” in ImageJ are visible in cyan. Scale bar = 20 µm.

(C) The second user-intervention when running the “identify nuclei” macro. The dialog box tells the user to check the ROIs and correct possible errors. The upper image shows a postmeiotic cyst in which nuclei have been automatically identified. The colored arrows point at ROIs that should be manually removed. Close-ups of the regions indicated by the colored arrows are displayed in the bottom row. The yellow arrow indicates two overlapping spermatid nuclei, which have been wrongly identified as one. The red arrow and the lower blue arrow indicate ROIs containing parts of cyst cell nuclei. The upper blue arrow indicates a spermatid nucleus that slightly overlaps with a cyst cell nucleus. Scale bars = 20 µm (upper image) and 2 µm (close ups lower row).

### Image analysis step 2: Measure DNA content

After having created an MIP and identified the nuclei, the DNA content in the nuclei can be measured. I created two macros to do so. The macro “Measure DNA content” is used to process one image at a time while the “Measure DNA content in subfolders” is used to batch processes all images in a folder and its subfolders. Images to be processed are recognized by the presence of an *“ cyst MIP.tif” image and its corresponding " nuclei ROIs.zip" file. T*he process of measuring the DNA content is the same with both macros. First, the MIP is loaded alongside the associated nuclei ROIs. If there are user defined “nuclei ROIs” they are discarded. The integrated pixel intensity of each nuclei ROI is measured. These values are then normalized to the mean of all measured nuclei. This resulting data is exported in table format with the extension “ DNA content.xls”. Thus, after running this macro, one new file will have been added to the folder of the MIP image: the results table.

### Data processing step 1: Merge DNA content results

After measuring the DNA content, the user is left with one result table per cyst that was analyzed. Ultimately, 10 to 20 cysts per genotype should be analyzed and, in most experiments, multiple genotypes are analyzed. To facilitate the handling of the data, all the results tables are merged into one data frame using the python script “merge DNA content results.py”. The user is asked to provide a folder which contains all the result files of one genotype. The result files can also be organized in subfolders. Data from the identified results files is added to a data frame with one row for each nucleus. To distinguish between nuclei from different cysts, the cysts are given a number and a column is added which contains the number of the cyst. The user is prompted to provide the genotype information for the batch of tables that were loaded. The genotype information is then added as a column in the dataframe. Then, another dialog window is created to ask whether the user wants to add results files from additional genotypes. Selecting yes reruns the first part of the code and adds the nuclei from the new genotype to the dataframe. Selection of “no” breaks the loop and lets the user select where to save the dataframe, as an .xslx table file.

### Data processing step 2: Calculate standard deviations of DNA content

The final readout of the assay is the standard deviation between nuclei in each cyst. The python script “process DNA content results.py” requires as input the combined results table that has just been generated and calculates the standard deviation between nuclei from the same cyst. The combined results table is loaded as a dataframe and the calculated standard deviation is added in a new column labeled “cyst\_stdev”. Furthermore, the average standard deviation between all cysts from one genotype is also calculated and added as a new column to the dataframe. Finally, the user can select where to save the dataframe as an .xslx table file. The compiled results table contains 11 columns and 1 row per analyzed nucleus (table 3.1). The ObjectNumber column indicates the number of the nucleus in the cyst. The Label column contains the name of the original image stack and the name of the nuclei ROI. The Area, Mean, IntDen (Integrated density) and RawIntDen (raw integrated density) have been measured in the nuclei ROIs. The RawIntDen has been normalized to the cyst average to obtain the norm signal (normalized signal). Further 2 columns give the cyst number and the genotype. The cyst\_stdev gives the standard deviation between the nuclei within a cyst. The Mean\_cyst\_stdev gives the standard deviation between the nuclei within a cyst averaged for all cyst within a genotype.

**Table 3.1. Illustrative DNA content analyses results table**

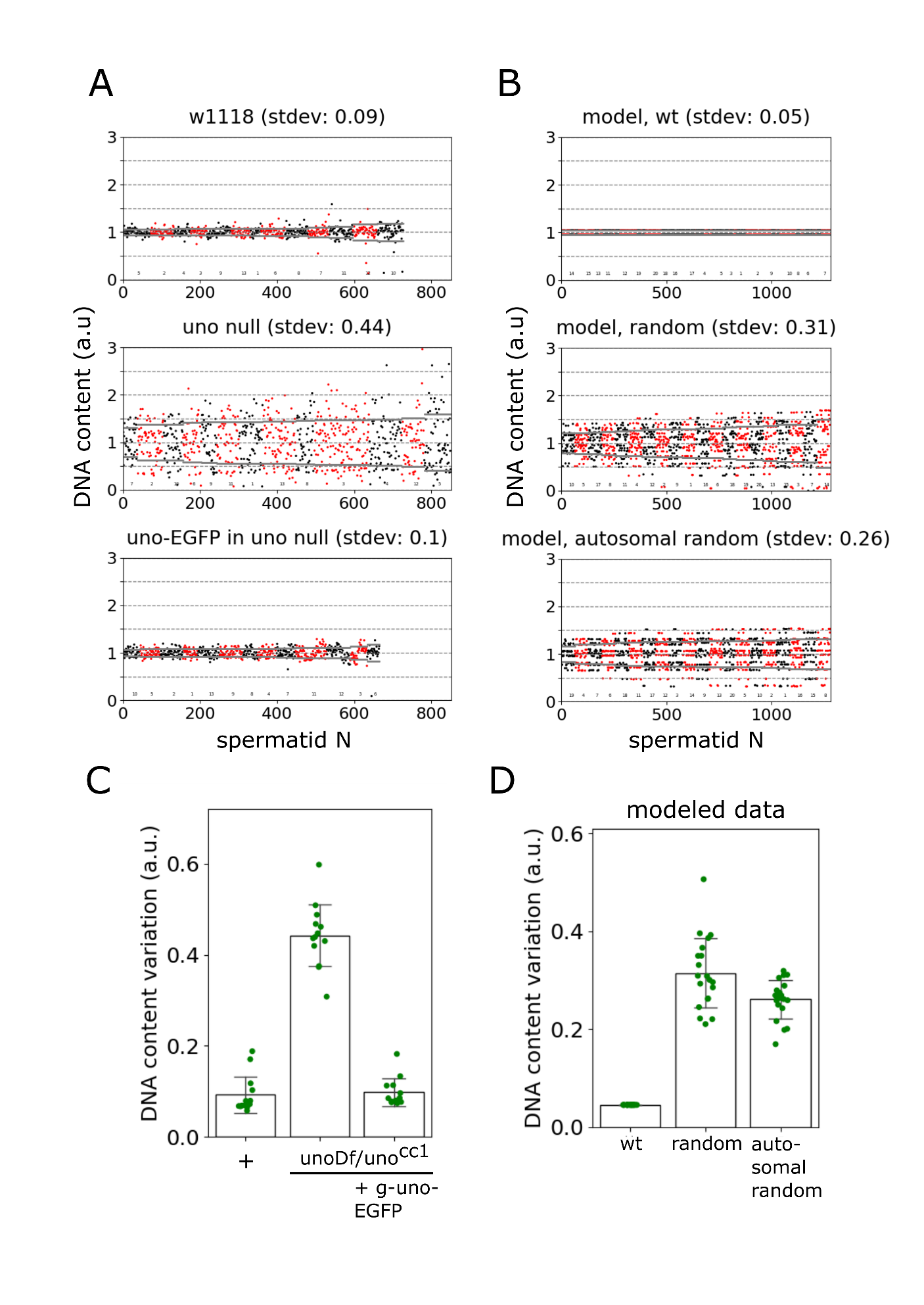
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Object**  **Number** | **Label** | **Area** | **Mean** | **IntDen** | **RawIntDen** | **norm signal** | **Cyst**  **Number** | **geno**  **type** | **cyst**  **stdev** | **Mean**  **cyst\_stdev** |
| 1 | Experi… | 12.35 | 12606 | 155691 | 5987741 | 1.27 | 1 | snm | 0.40 | 0.45 |
| 2 | Experi… | 7.12 | 16773 | 119495 | 4595686 | 0.98 | 1 | snm | 0.40 | 0.45 |
| 3 | Experi… | 4.84 | 8731 | 42226 | 1623995 | 0.35 | 1 | snm | 0.40 | 0.45 |
| 4 | Experi… | 4.24 | 13514 | 57275 | 2202747 | 0.47 | 1 | snm | 0.40 | 0.45 |
| 5\* | Experi… | 9.83 | 14659 | 144080 | 5541185 | 1.18 | 1 | snm | 0.40 | 0.45 |

\*This table is for illustrative purposes and therefore contains only 5 rows.

### Data visualization on scatterplot and barplot

To visualize the results of the DNA content analyses I provide two options. The first option through “scatterplot DNA content results.py” creates a series of scatterplots where each nucleus is represented as a datapoint spread across the x axis. The y axis gives the normalized DNA content (Fig 3.3A). One scatterplot is drawn per genotype. The nuclei belonging to different cysts are visually distinguishable through alternating black and red colors. The standard deviation of the nuclei in each cyst is shown as a gray bar. The average standard deviation of the cysts is given in the plot title. The cysts are ordered from lowest to highest standard deviation. To allow trace-back to the original cyst/images the image numbers are given below the datapoints belonging to a certain image. The high information content of these scatterplots means that they are very useful to assess the quality of the data and identify outliers.

The second option for the visualization of DNA content analyses is through “barplot DNA content results.py”. In the barplot, the cysts are treated as main elements (Fig 3.3C). The x axis indicates the genotype. The y axis gives the standard deviation between the nuclei in a cyst. The bars represent the mean standard deviation of all the cyst in one genotype. The error bars represent the standard deviation of the standard deviation inside the cysts of that genotype. The advantage of the barplot is that different genotypes can be directly compared on one plot. As the standard deviation of the nuclei within a cyst is the main readout of the assay, this plot is ideal to visualize the results.

 **Figure 3.3: Visualizing results from DNA content analyses.**

*(A) Scatter plots representing the results of a DNA content assay. The plots were created with the* “scatterplot DNA content results.py” script. See text for information on plot elements. The plotted data is from part 1 of this thesis. The genotypes correspond to the genotypes in the publication indicated as unoDf/unocc1 (for uno null) and unoDf/unocc1 + g\_uno-EGFP (for uno-EGFP in uno null).

*(B) Scatter plots displaying the DNA content variability generated by modelling regular as well as different types of irregular meiotic chromosome segregation. The modeled data was obtained with the “model DNA content.py” script. The plots were created through the* “scatterplot DNA content results.py” script. The title of each plot indicates the modeled type of meiotic chromosome segregation: wt for regular segregation during meiosis I and II, random for random segregation during meiosis I followed by regular segregation during meiosis II, and autosomal random for regular XY segregation and random autosome segregation during meiosis I followed by regular segregation during meiosis II. See the text for information on plot elements and data modeling.

*(C) Results of a DNA content assay visualized with a bar plot. The plot was created with the* “barplot DNA content results.py” script. The genotype labels were edited manually in Inkscape for improved readability. See text for information on plot elements. The plotted data is from part 1 of this thesis.

*(D) Results of a modeled in silico DNA content assay visualized on a bar plot. The modeled data was obtained with the “model DNA content.py” script. The plot was created with the* “barplot DNA content results.py” script. The different types of modeled chromosome segregation (wt, random and autosomal random; see above (B)) are indicated below the bars. See text for information on plot elements and data modeling.

### Discussion of the accuracy of the DNA content analysis assay

An early version of the procedure for measuring the DNA content in postmeiotic spermatids has been used to investigate the temporal dynamics of *mnm* and *snm* requirements during spermatocyte maturation (Sun *et al.*, 2019). The final version has been used for the characterization of *uno* (see part 1 of this thesis). Besides these published results, the final procedures has been used repeatedly in the lab to evaluate transgene functionality in rescue experiments.

#### Distinguishing between normal and random segregation in meiosis I

The most basic test for the new assay was to check whether it can reliably distinguish between regular wild-type and abnormal random segregation of chromosomes during meiosis I. Random segregation of chromosomes in meiosis I is known to occur in *snm* and *mnm* mutants based on analyses with a genetic 4th chromosome non-disjunction assay and by XY FISH (Thomas *et al.*, 2005; Sun *et al.*, 2019). The new DNA content analysis assay was found to distinguish comfortably between wild type and *mnm* or *snm* mutants (Sun *et al.*, 2019). Similarly, the difference between wild-type and *uno* mutants observed by time lapse imaging was clearly confirmed by the DNA content analyses (Part 1 of this thesis Fig 2C). In the analyses performed for the characterization of *uno*, the standard deviation within a wild-type cyst was 0.1 on average. The standard deviation of the standard deviations observed in different wild-type cysts was 0.037. In contrast, the standard deviation within a *uno* mutant cyst was 0.46 on average, and the standard deviation of the standard deviations observed in different *uno* mutant cysts was 0.069. A t-test revealed a highly significant difference in the postmeiotic DNA content variability between wild type and *uno* mutants (P-value 1.18 \* 10-13). Therefore, the novel assay is clearly capable of detecting differences between normal and completely random segregation of chromosomes in meiosis I.

#### Comparing measured and modeled data to assess the sensitivity of the assay

To what extent might the assay also allow for an accurate detection of smaller differences in the error rates of chromosome segregation? In general, the observed variation in the standard deviation of the DNA contents per nucleus measured in different cysts of the same genotype was relatively high. In principle, this observed variation might reflect technical and/or biological variability. To explore the theoretically predicted variability both within and between cysts in case of random segregation, I created a model for chromosome segregation in meiosis in python (S 3.11,“model DNA content.py”). Three types of chromosome segregation in meiosis I were modeled: normal segregation, complete random segregation as in *mnm, snm* or *uno* mutants, as well as complete random segregation of autosomes but normal segregation of sex chromosomes as in *teflon* mutants. The models first distribute eight chromosome objects as in meiosis I into two arrays either equationally, randomly, or randomly except for X and Y. These arrays are then duplicated to obtain a pair of spermatids, mimicking normal meiosis II. 16 of these spermatid quartets are united, creating a 64-spermatid array object, the cyst equivalent. The amount of DNA in each *in silico* spermatid is calculated based on which chromosomes it contains. For all chromosomes except the Y chromosome the DNA content per chromosome is based on the number of basepairs in the NCBI genome browser (release 5). For the heterochromatic Y chromosome the DNA content has been estimated at 40 Mb (Carvalho, 2002). It must be noted that the reference genome does not include most of the pericentromeric heterochromatin and thus the relative size of the Y chromosome is overrepresented by about 10% compared to the other chromosomes. This is small enough to not affect the overall results of the modeling. To mimic the fact that the DNA content is measured with a 2D maximum intensity projection of the DNA signals, and not by integrating DNA signals in 3D, I converted the DNA content of *in silico* spermatid nuclei according to the formula that describes the relation between the volume of a sphere and the area of the largest equatorial section (A=π1/3(6V)2/3). The values of the transformed DNA content of the *in silico* spermatid nuclei were then normalized to the average of the *in silico* cysts with 64 spermatids. This was repeated 20 times for each type of meiosis (see above) and the data was exported as .xlsx file which was used for plotting using the scripts “Scatterplot DNA content results” and “Barplot DNA content results” described above (Fig 3.3B, 3.3D). In the model with normal segregation, the spermatids can only have two possible values of DNA, depending on whether they have an X or a Y chromosome. The standard deviation after this regular distribution resulting in the *in silico* model was 0.045 in every cyst. As expected, random segregation in MI resulted in a more complex distribution in the *in silico* model. The average standard deviation from the 20 modeled cysts was 0.31, with values for individual cysts ranging from 0.21 to 0.56. After random segregation of autosomes and regular sex chromosome segregation, the *in silico* model generated an average standard deviation of 0.26, with values for individual cysts ranging from 0.17 to 0.39. The modeling was also used for analysis of far greater cyst numbers. By generating 1000 cysts per type of meiosis, the random segregation model resulted in an average of the standard deviations across all cysts of 0.30, and the intra-cyst standard deviation was on average 0.056. In case of autosomal random segregation with regular sex chromosome segregation, the corresponding numbers were 0.27 and 0.045, respectively.

#### Estimating the variation between nuclei in a cyst

The comparison between measured and modeled intra-cyst variation can serve as an estimate of the technical variation. For completely regular segregation, the measured intra-cyst standard deviation is 0.1 while the modeled value is 0.045. In case of AHC mutants, the measured intra-cyst standard deviations are 0.41 (*mnm*), 0.45 (*snm*) and 0.46 (*uno*) (Sun *et al.*, 2019)(part 1 of this thesis), while the modeled value is 0.31. These difference between measured and modeled values are substantial but not identical in the two different types of meiosis I. Two possibilities might be considered with regard to the variable extent of the difference between measured and modeled: either the technical variability increases when the biological variation increases, or the model underestimates the measured biological variation. Fortunately, in both scenarios the value of intra-cyst standard deviation would mainly depend on the biological variation, ensuring the robustness of the assay.

#### Estimating the variation between cysts

A more important comparison concerns the differences between modeled and measured inter-cyst variation. In the modeled wildtype segregation, all cysts are equal and therefore the inter-cyst standard deviation is 0. In contrast, the corresponding measured value is 0.045. In this case we can be confident that all of the measured variation comes from technical variation. The modeled inter-cyst standard deviation in case of completely random segregation is 0.056, while the measured values were 0.043 (*snm*), 0.039 (*mnm*), and 0.069 (*uno*) (Sun *et al.*, 2019)(part 1 of this thesis). The modeled value of 0.056 does not contain technical variation, while the measured values do. To better compare both values, it is possible to add the technical variation measured in the wildtype to the modeled value based on complete random segregation. To do so the square of the measured inter-cyst standard deviations measured for the wildtype is summed with the square of the inter-cyst standard deviation predicted *in silico* after random segregation. The square root of the resulting sum yields the *in silico* value for random segregation converted to include the technical variation. It was found to be 0.072. This modeled inter-cyst standard deviation of 0.072 can now be compared to the measured values of 0.043 (*snm*), 0.039 (*mnm*), and 0.069 (*uno*). Given that the modeled value is higher than the measured values it seems that the DNA content analyses adds even less technical variability at the level of inter-cyst variation in case of random segregation than in case of normal segregation. Thus, the measured inter-cyst standard deviation in case of random segregation is mainly based on biological variation and not on technical variation.

#### Discussion

The comparison between the modeled and measured distributions suggests that the assay overestimated the variation between nuclei in a cyst. On the other hand, the variation between cyst might be slightly underestimated. Finding potential explanations for these observations is difficult. It is expected that the simple assumption made by the model do not reflect the complexity of biology. One assumption of the model is that the relation between the true DNA content and the measured DNA content is the same as between a volume of a sphere to the area of the equatorial disc. I presume the true relation between these two to be more complex. Another potentially incorrect assumption is that of a total random segregation in AHC mutants. It is conceivable that highly unbalanced divisions (e.g. 1 to 7 and 0 to 8) are less likely than predicted. Possible biological reasons for this could be that a daughter cell with 8 chromosomes would be simply too crowded, or that an accumulation of chromosomes on one pole increases the chance that microtubules from the other pole pull on the remaining chromosomes. The assumption that meiosis II is completely regular in AHC mutants is likely also partially incorrect. Missegregation of univalents during meiosis I can lead to the formation of DNA bridges which can interfere with successful cytokinesis and lead to multipolar meiosis II. The main conclusion from the modeling is that the measured inter-cyst variation, given by the standard deviation of the standard deviation of the nuclei in a cyst, is likely either correctly estimated or slightly underestimated by the assay. This means the robustness of the assay is not compromised by a high technical variation between cysts. Furthermore, I expect the modeled data combined with the DNA content analyses performed so far to be useful to plan future experiments.

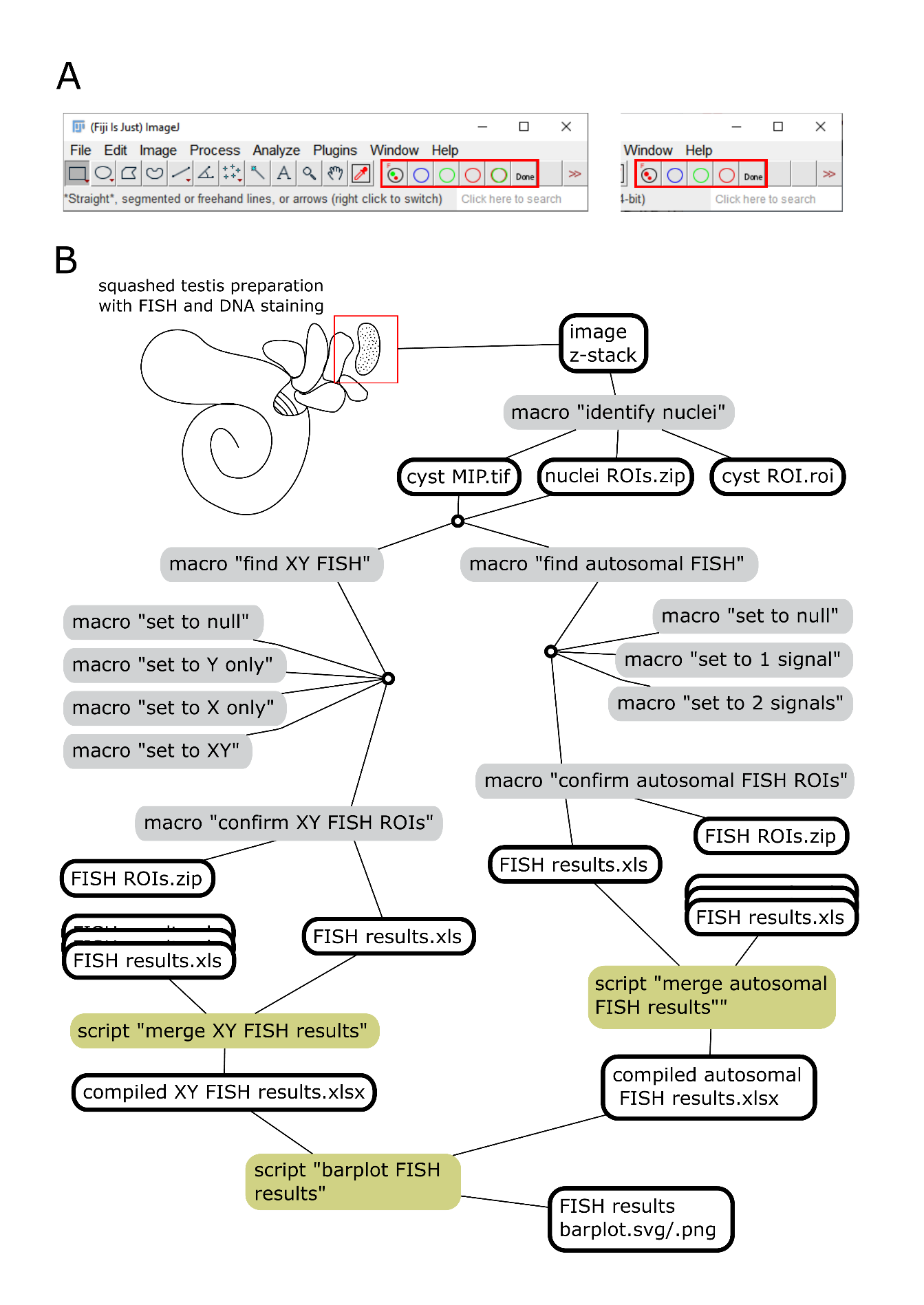
### Conclusion

Overall, I conclude that the DNA content analysis assay is performing as intended. The strongest point of this assay is its efficiency. Testis squash preparation slides can be prepared and imaged within a few hours and the image and data analyses pipeline takes only a few minutes. With the provided macros and scripts, it is technically easy to perform. The assay is robust and quantitative. The weakest point of the assay is its limited specificity. From the results of the assay one cannot conclude whether segregation is random in meiosis I or in meiosis II. However, in many experimental setups, such as rescue experiments, the phenotype of a certain mutant is already known. Moreover, the assay can be coupled with live imaging or FISH analyses to confirm which meiotic division is affected.

## **Method 2:** Pipeline to analyze chromosome segregation defects using FISH

### Overview

The assay for chromosome segregation defects based on DNA content analysis can be enhanced by performing a FISH assay in parallel. An XY FISH assay will provide a second read-out of chromosome segregation defects and add some information on whether chromosomes segregate randomly in meiosis I or in meiosis II. Random segregation in meiosis I results in some spermatids having an X and a Y chromosome, while random segregation in meiosis II results in some spermatids having two X or two Y chromosomes. If segregation in both divisions is random, then both types of aneuploidy can be detected. In addition to this, coupling the DNA content assay to the XY FISH assay provides information on whether there is a difference between segregation of sex chromosomes and autosomes, like in teflon mutants where autosomes segregate randomly but sex chromosomes segregate normally (Tomkiel, Wakimoto and Briscoe, 2001). A FISH assay is time-consuming for two reasons. First, the preparation of the microscopy slides is a two-day process. Second, FISH signals are manually assessed in each spermatid. Usually about 640 spermatocytes (i.e. 10 cysts) are assessed per genotype. To speed up this step of image analyses, I created an image analyses pipeline to assist in analyzing chromosomal FISH staining in spermatids (XY FISH or autosomal; Fig 3.4A, 3.4B, 3.5A). The pipeline is compatible with the DNA content analysis pipeline and shares the same process of nuclei identification. For the steps of identifying FISH signals, I coded ImageJ scripts. I also coded python scripts to process the resulting data and visualize it. This image analysis pipeline can be performed without any knowledge of programming. The requirements to run the pipeline are the same as for the DNA content analysis. In this case the ImageJ toolsets that need to be installed are the “XY FISH analyses.txt” and “autosomal FISH analyses.txt”. The code is available as supporting information S3.2, S3.3, S3.8, S3.9, S3.10 and on github (https://github.com/BioJoe/automated-FISH-analyses). For slide preparations and microscopy also see above (DNA content analysis).



**Figure 3.4: Pipeline to assay chromosome segregation defect based on FISH staining.**

(A) Image of the main window of FIJI with the FISH analyses toolset loaded, either the XY FISH analyses toolset (left) or the autosomal FISH analyses toolset (right). The individual macros of the toolset can be run by clicking on the icons highlighted with the red rectangle. The first icon with “F” runs the “find XY FISH” or “find autosomal FISH” macro. The colored circle icons are used to run the macros to correct the FISH identification. The last icon “Done” is used to run the “confirm XY FISH ROIs” or “confirm autosomal FISH ROIs” macro.

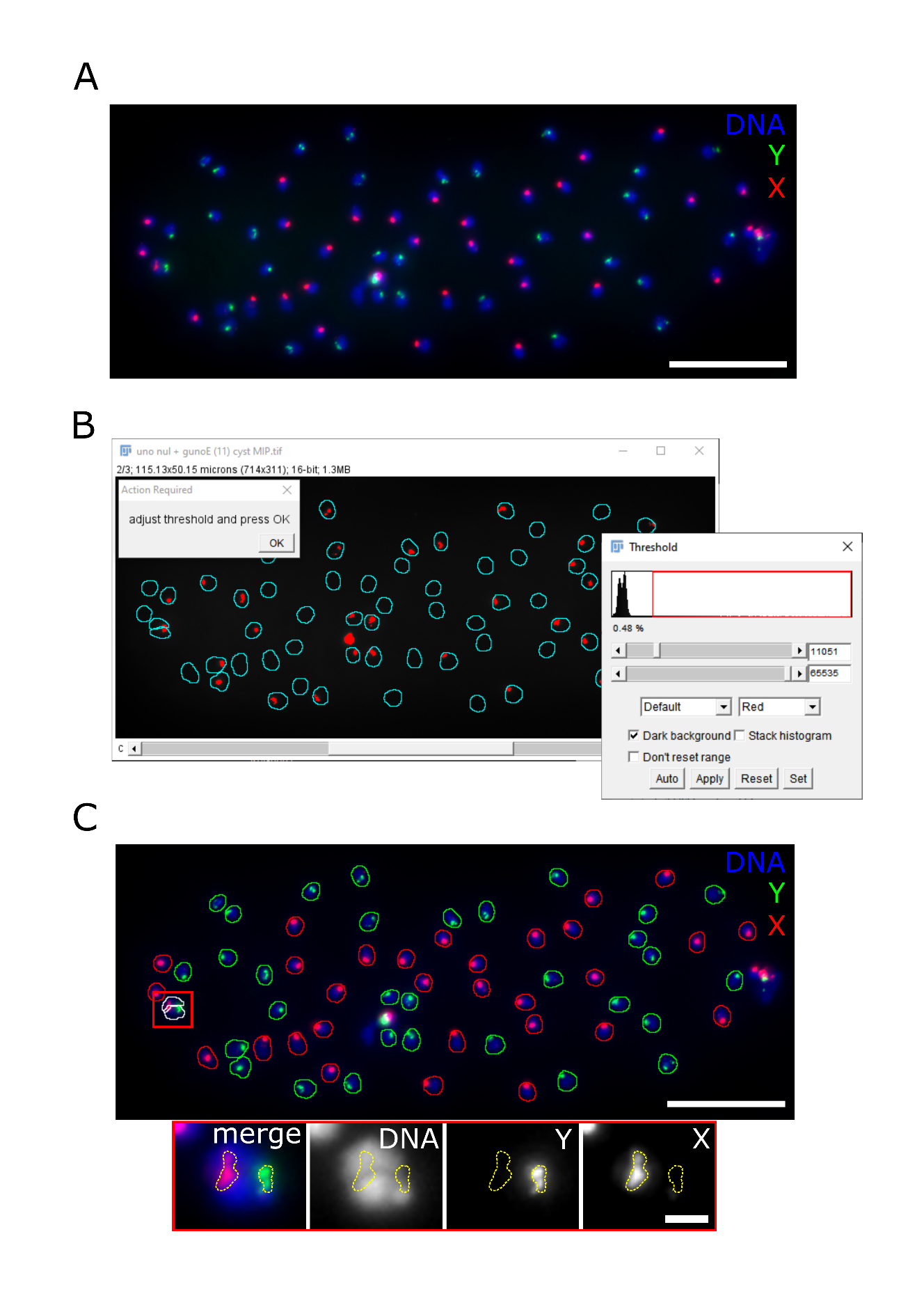
(B) Schematic representation of the FISH analyses pipelines. The drawing represents an adult testis with cysts spilling out at a cut site. Files created throughout the pipelines are represented by their extension name and a black outline. ImageJ macro have a grey background box. Python scripts have a yellow-green background box. The lines going from a macro or script to a file indicate that this file is created by running this macro or script. The lines going from a file to a macro or script indicate that this file is required to run the macro or script.

Image analyses XY FISH

This pipeline to analyse XY FISH requires the “cyst MIP.tif” and “nuclei ROIs.zip” files created in the first step of the DNA content analyses (“Identify nuclei”). After that, the pipeline starts by running the “Identify XY FISH signals” macro which first asks the user for the location of an MIP image. The corresponding MIP and nuclei ROI files are loaded. To improve detection of FISH signals at the border of nuclei the nuclei ROIs are enlarged by 1 pixel. The user is then asked to set a threshold for detection of the Y signal in the green channel (Fig 3.5B). If the quality of the FISH staining is good, it is easy to find a threshold which identifies nearly all signals correctly and does not add false positives. The code then checks for the presence of green signal above background in each of the nuclei ROIs. If a cluster of more than 10 pixels with green signal is found the color of the nuclei ROI is changed to green. The code is then repeated for the identification of X FISH signals in the red channel. If an X FISH signal is found in a nucleus which had no Y FISH signal this nucleus ROI is colored red. If the nucleus contains both an Y and X signal the nucleus ROI is colored white. If no signal is detected, the nucleus ROI stays blue.

At this point the user is the left with the MIP and an overlay of colored nuclei ROIs (Fig 3.5C). The user should now check whether the FISH signals have been correctly identified and correct eventual mistakes. A set of 4 simple macros can be used to change the assigned ROI color. Blue is for nuclei without FISH signal, green for those with only a Y FISH signal, red for only X FISH and white for Y and X signals in one nucleus. The macro names are: “set to null”, “set to Y only”, “set to X only”, *“*set to XY”). In the example cyst analyzed here, two nuclei had to be manually corrected (Fig 3.5C). These 2 nuclei were partially overlapping, and the X FISH signal was identified in both ROIs instead of only the upper one. The Y FISH signal was identified in the upper ROI although to my assessment it belongs to the lower nucleus. Most green signal is in the lower ROI, but it is below the threshold and thus not recognized. In general, less than 5% of the nuclei need to be manually corrected.

When the assessment of the FISH signals in all nuclei ROIs has been verified, the macro “confirm XY FISH ROIs” is used to save the ROIs and export the results. The number of ROIs of each color are counted and saved in table format with the extension “ FISH results.txt”. The FISH ROIs are saved with extension " FISH ROIs.zip" (Fig 3.4).



**Figure 3.5: Semi-automated identification of FISH signals in spermatid nuclei.**

(A) Maximum intensity projection of an image stack containing an early postmeiotic cyst from a testis squash preparation labeled with the DNA stain Hoechst 33258 and by FISH probes for the X (red) and the Y (green) chromosome. In this MIP from a wild-type testis, each nucleus contains either a green (Y) or a red (X) signal as expected after regular sex chromosome segregation during meiosis I. In AHC mutant cysts, on average half of the nuclei contain no sex chromosome (no FISH signals) or both (red and green FISH signals). Scale bar = 20 µm.

(B) First user-intervention when running the “find XY FISH” or “find autosomal FISH” macro. The dialog box in the left upper corner tells the user to adjust the threshold to detect FISH signals. The image shows the nuclei ROIs in cyan and in red the pixels above threshold in real time. The window on the right shows the inbuilt thresholding tool from ImageJ used to adjust the threshold.

(C) The upper image shows a cyst after the FISH signals have been identified using the “find XY FISH” macro. The nuclei in which a Y FISH signal has been identified have a green outline and the nuclei in which an X FISH signal has been identified have a red outline. Nuclei in which both FISH signals have been identified have a white outline. Not present in this example are nuclei in which no FISH signals have been identified, they would have a blue outline. The red rectangle indicates two nuclei in which the FISH signals have been incorrectly identified, requiring manual correction. Close ups of these nuclei are shown in the bottom row. The Y and X signals have been outlined with a yellow dashed line. For these two nuclei, FISH signals cannot be attributed to either nucleus with absolute confidence.

### Image analyses autosomal FISH

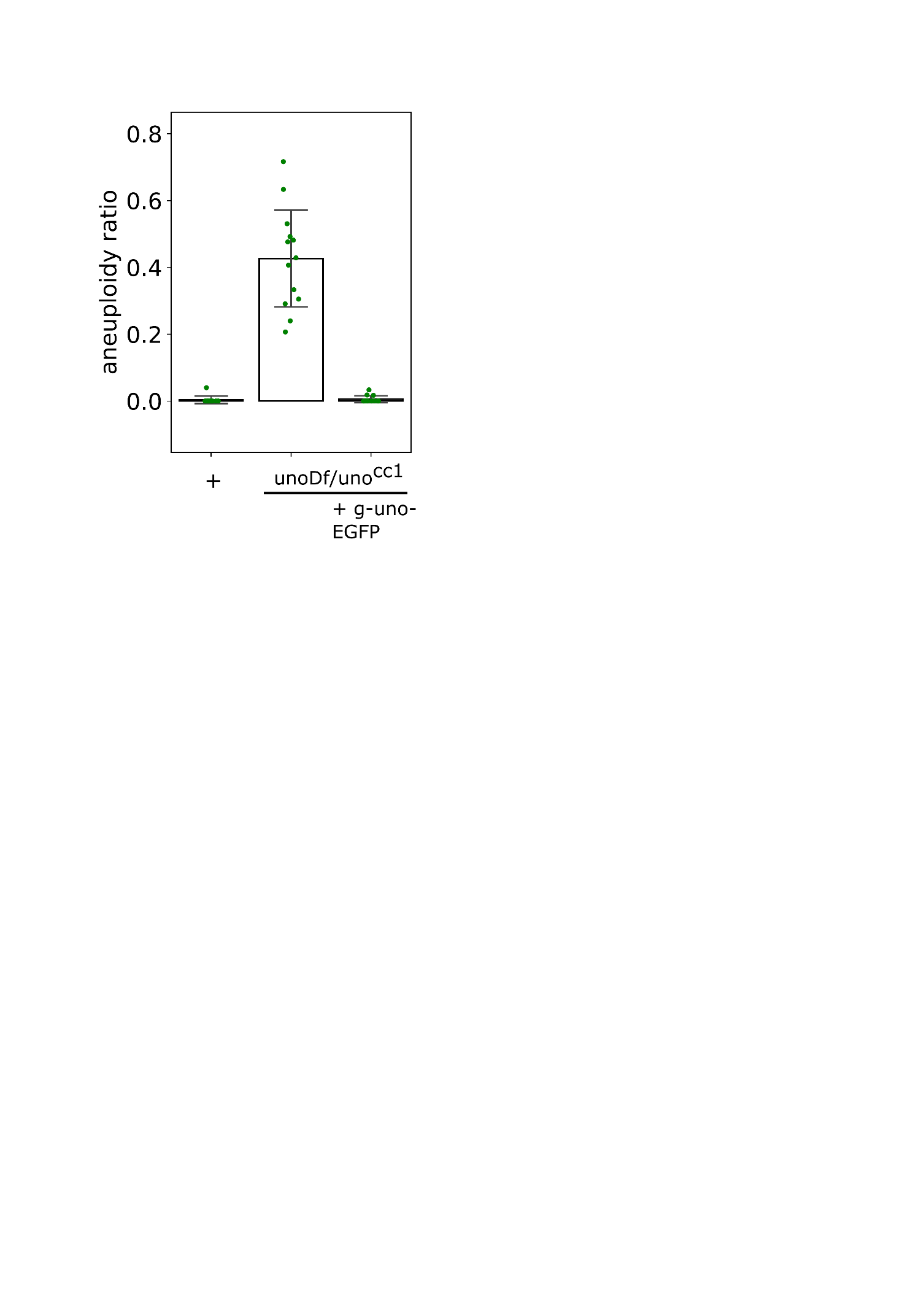
The toolset to analyze autosomal FISH is a modified version of the toolset to analyze XY FISH. The requirements are the same and most of the code is shared. The main difference lies in how the FISH signals are identified. In the XY FISH image analyses each chromosome is identified in a different color channel. In contrast, in the autosomal FISH image analyses signals from both chromosomes need to be identified in the same color channel. In the macro “Identify autosomal FISH signals”, after setting the signal threshold, the code checks not only whether there is signal above background in a nuclei ROI, but also whether there are one or two patches of signal. If one patch of more than 10 pixels with signal is found the color of the nuclei ROI is changed to green. If two such patches are found within the same ROI, the outline color of the ROI is changed to red. If no signal patch is detected within a ROI, the nucleus ROI becomes blue. At this point the user is the left with the MIP and an overlay of colored nuclei ROIs. The following 3 simple tools are used to manually correct eventual errors. They simply change the color of the selected ROI to either blue, green, or red (set to null, set to 1 signal, set to 2 signals). When the assessment of the FISH signals in all nuclei ROIs has been verified the macro “confirm autosomal FISH ROIs” is used to save the ROIs and export the results. The number of ROIs of each color are counted and saved in table format with the extension “ FISH results.txt”. The FISH ROIs are saved with extension " FISH ROIs.zip" (Fig 3.4).

### Data processing

To merge the result files from FISH analyzes I created two scripts, one for each type of FISH analyses (“merge XY FISH results.py”, “merge autosomal FISH results.py”). The scripts share a lot of code but differ in how the data is extracted from the result files and how the aneuploidy rate is calculated (see below). Both scripts ask the user for a directory containing the results files of one analyzed genotype and scan that directory and its subdirectories for relevant result files. From each of the identified files the data is extracted and placed as a row in a dataframe. The dataframe also adds the directory and name of analyzed image as columns. The user is then asked to provide the genotype information of the loaded results files that is added as an additional column. A dialog window is created to ask whether the user wants to add results files from another genotype. Selecting “yes” re-runs the first part of the code and adds the new rows to the dataframe. Selection “no” breaks the loop and continues with the calculation of the ratio of aneuploid nuclei for each cyst, which is added as a column. For XY FISH this ratio is (2\* number of XY nuclei/sum of X, Y and 2\* XY nuclei). For autosomal FISH this ratio is (2\* number of nuclei with double signal/sum of single signal and 2\*double signals). The user can then select where to save the final dataframe as an .xslx table file.

### Data visualization

The results of FISH assays for chromosome segregation defects are ideal to represent in the form of a barplot. This is done by running the script “barplot FISH results.py”. This script extracts the genotype and aneuploidy ratio for each cyst in the compiled FISH results file (XY or autosomal). A barplot is then created with the genotype on the x axis and the aneuploidy ratio on the y axis (Fig 3.6). The plot includes errorbars with the standard deviation. The aneuploidy ratio for cysts is also displayed as single datapoints. The plot can be saved as raster image (.png) or as vector image (.svg).

**Figure 6: Results of an XY FISH assay visualized on a barplot.**

The plot was created with the “barplot FISH results.py” script. The genotype labels were manually edited in Inkscape for improved readability. See text for information on plot elements. The plotted data is from part 1 of this thesis.

## Protocol for chromosome segregation analyses (DNA content and FISH)

1. Prepare microscopy slides with fixed squashed testes with DNA stain (see White-Cooper, 2004) according to protocol 3.3.2. Optional: perform FISH staining as previously described (Sun et al., 2019).
2. Take 40x image stacks of post meiotic cysts with early round spermatids with 280 nm z-spacing. Image 10-20 cysts per genotype.
3. Load the “DNA content analyses toolset” in FIJI
4. Click on the icon with the I to run the “Identify nuclei” macro and follow the instructions. Delete ROIs on cyst nuclei when prompted to do so. If two overlapping nuclei are recognized as one also delete this ROI. If FISH analyzes is intended draw the correct ROIs manually and press t to add to the ROI manager.
5. Repeat step 4 until all images have been processed.
6. Run the “merge DNA content results” python script and follow the instructions to add results from multiple genotypes.
7. Run the “process DNA content results” python script.
8. Run the “scatterplot DNA content results” python script. Review the data on the plot. Genotypes with wildtype chromosome segregation have around 0.1 average standard deviation. Genotypes with random chromosome segregation in meiosis I have around 0.42 average standard deviation. Save the plot as .svg and/or .png.
9. To better compare genotypes in one plot run the “barplot DNA content results” python script. Save the plot as .svg and/or .png.

The following steps are only needed if FISH signals are analyzed

1. Load the toolset “XY FISH analyses” or “autosomal FISH analyses”
2. Click on the icon with the F to run the “Identify … FISH signals” macro and follow the instructions.
3. Look at each nucleus and make sure the signals have been identified correctly. To correct eventual mistakes, use the tools with the colored circle icons. For XY FISH: blue is for no signal, green is for Y only, red is for X only and white is for X and Y. For autosomal FISH: blue is for no signal, green is for 1 signal, red is for 2 signals.
4. When all nuclei have been assigned the correct color click the “Done” icon to run the “confirm … FISH ROIs” macro.
5. Run the “merge … FISH results” python script and follow the instructions to add results from multiple genotypes.
6. Run the “barplot FISH results” python script to obtain a barplot comparing the different genotypes. Save the plot as .svg and/or .png.

## References

Carvalho, A. B. (2002) ‘Origin and evolution of the Drosophila Y chromosome’, *Current Opinion in Genetics and Development*. doi: 10.1016/S0959-437X(02)00356-8.

Chaurasia, S. and Lehner, C. F. (2018) ‘Dynamics and control of sister kinetochore behavior during the meiotic divisions in Drosophila spermatocytes’, *PLoS Genetics*. doi: 10.1371/journal.pgen.1007372.

Grell, E. H. (1963) ‘Distributive Pairing of Compound Chromosomes in Females of Drosophila Melanogaster.’, *Genetics*.

Hart, T. A. *et al.* (2008) ‘Condensin II resolves chromosomal associations to enable anaphase I segregation in Drosophila male meiosis’, *PLoS Genetics*. doi: 10.1371/journal.pgen.1000228.

McKee, B. D. and Karpen, G. H. (1990) ‘Drosophila ribosomal RNA genes function as an X-Y pairing site during male meiosis’, *Cell*. doi: 10.1016/0092-8674(90)90215-Z.

McKee, B. D., Yan, R. and Tsai, J.-H. (2012) ‘Meiosis in male Drosophila’, *Spermatogenesis*. Informa UK Limited, 2(3), pp. 167–184. doi: 10.4161/spmg.21800.

Schindelin, J. *et al.* (2012) ‘Fiji: An open-source platform for biological-image analysis’, *Nature Methods*. doi: 10.1038/nmeth.2019.

Sun, M. S. *et al.* (2019) ‘Mnm and snm maintain but do not establish achiasmate homolog conjunction during drosophila male meiosis’, *PLoS Genetics*. doi: 10.1371/journal.pgen.1008162.

Thomas, S. E. *et al.* (2005) ‘Identification of Two Proteins Required for Conjunction and Regular Segregation of Achiasmate Homologs in Drosophila Male Meiosis’, *Cell*, 123(4), pp. 555–568. doi: http://dx.doi.org/10.1016/j.cell.2005.08.043.

Tomkiel, J. E., Wakimoto, B. T. and Briscoe, A. (2001) ‘The teflon gene is required for maintenance of autosomal homolog pairing at meiosis I in male Drosophila melanogaster.’, *Genetics*.

Tsai, J. H., Yan, R. and McKee, B. D. (2011) ‘Homolog pairing and sister chromatid cohesion in heterochromatin in Drosophila male meiosis I’, *Chromosoma*, 120(4), pp. 335–351. doi: 10.1007/s00412-011-0314-0.

Wakimoto, B. T., Lindsley, D. L. and Herrera, C. (2004) ‘Toward a comprehensive genetic analysis of male fertility in Drosophila melanogaster’, *Genetics*. doi: 10.1534/genetics.167.1.207.

White-Cooper, H. (2004) ‘Spermatogenesis: analysis of meiosis and morphogenesis.’, *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1385/1-59259-665-7:45.