

**Methods and Genome-Wide Association Study for Meiotic Nondisjunction of Chromosome  
21**

by

**Jonathan McKelvey Chernus**

BS, BA, University of Pittsburgh, 2010

MA, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of the  
Graduate School of Public Health in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH  
GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

**Jonathan McKelvey Chernus**

It was defended on

July 27, 2021

and approved by

**Dissertation Advisor:**

Eleanor Feingold, PhD

*Professor*

*Departments of Biostatistics and Human Genetics*

*Graduate School of Public Health*

*University of Pittsburgh*

**Committee Members:**

Candace M. Kammerer, PhD

*Associate Professor*

*Department of Human Genetics*

*Graduate School of Public Health*

*University of Pittsburgh*

John R. Shaffer, PhD

*Assistant Professor*

*Departments of Human Genetics and Oral and Craniofacial Sciences*

*Graduate School of Public Health and School of Dental Medicine*

*University of Pittsburgh*

George C. Tseng, ScD

*Professor*

*Departments of Biostatistics, Human Genetics, and Computational & Systems Biology*

*Graduate School of Public Health*

*University of Pittsburgh*

Copyright © by Jonathan McKelvey Chernus

2021

# **Methods and Genome-Wide Association Study for Meiotic Nondisjunction of Chromosome 21**

Jonathan McKelvey Chernus, PhD

University of Pittsburgh, 2021

## **ABSTRACT**

Up to one quarter of human conceptions may be aneuploid, having too many or too few chromosomes relative to the standard 23 pairs. Most often this results from nondisjunction in maternal meiosis, making such errors a leading cause of pregnancy loss and congenital abnormalities. Prior research has established advanced age and altered patterns of meiotic recombination as risk factors for maternal meiotic nondisjunction and has shown that meiosis I and II errors may involve different mechanisms, but genetic risk factors have not been systematically investigated. The goal of this dissertation is to advance our knowledge of aneuploidy by identifying and characterizing common genetic variants associated with maternal meiotic nondisjunction of chromosome 21, the most common aneuploid condition in conceptions that survive to term.

The first aim was to perform a candidate gene and genome-wide association study (GWAS) in which cases are mothers who have had a child with Down syndrome and controls are the fathers. We found plausible associations at variants at relevant loci. Stratifying by the stage of meiosis in which nondisjunction occurred (MI or MII), our results are consistent with general nondisjunction risk factors as well as some that could be specific for MI or MII.

In the second aim, we called recombination events on chromosome 21 in our data set in order to classify cases (mothers) according to their recombination profiles. We therefore developed and implemented novel methods for calling recombination events in both trios and dyads, finding

that full-data trios can be used to successfully train the method for calling recombination in dyads, which contain less information.

The third aim was to further characterize the candidate gene and GWAS associations by performing stratified analyses in subgroups of mothers defined by recombination profile and maternal age. We interpret the associations in the context of possible meiotic error mechanisms.

The public health significance of this research is its improvement of our understanding of the genetic architecture of meiotic errors, a leading factor in pregnancy loss and congenital defects. Eventually this could lead to identifying those at higher risk of meiotic errors and enabling more informed reproductive choices.

**Table of Contents**

**1.0 INTRODUCTION..... 1**

**1.1 BACKGROUND AND MOTIVATION..... 1**

**1.1.1 Meiosis, Aneuploidy, and Established Risk Factors .....1**

**1.1.2 Possible Mechanisms and Genes of Interest .....2**

**1.2 GENOTYPING ARRAY DATA IN THE CONTEXT OF DOWN SYNDROME... 4**

**1.2.1 Family Data and Informative Markers.....4**

**1.2.1.1 Methods for Trios ..... 5**

**1.2.1.2 Methods for Dyads..... 9**

**1.2.2 Description of Data Set .....12**

**1.3 AIMS AND SUMMARY OF THIS DISSERTATION ..... 13**

**1.3.1 Finding Genetic Risk Factors for Nondisjunction .....13**

**1.3.2 Using Family Data to Call Recombination Events on Chromosome 21.....14**

**1.3.3 Characterizing Genetic Risk Factors for Nondisjunction.....14**

**2.0 A CANDIDATE GENE ANALYSIS AND GWAS FOR GENES ASSOCIATED WITH MATERNAL NONDISJUNCTION OF CHROMOSOME 21 ..... 15**

**2.1 CHAPTER OVERVIEW..... 16**

**2.2 AUTHOR SUMMARY ..... 16**

**2.3 INTRODUCTION ..... 17**

**2.4 METHODS..... 22**

**2.4.1 Study Sample.....22**

**2.4.1.1 Ethics Statement ..... 22**

2.4.1.2 Genotyping .....	23
2.4.1.3 Adjustment for Population Structure .....	25
2.4.1.4 Imputation .....	25
2.4.1.5 Phenotyping.....	25
2.4.2 Analysis .....	27
2.4.2.1 Sample Size.....	27
2.4.2.2 Association Studies .....	27
2.4.2.3 Maternal Age Effect.....	29
2.4.2.4 Candidate Gene Analyses.....	30
2.4.2.5 Follow-up Analyses to Examine Top-Ranked GWAS Signals.....	30
2.5 RESULTS.....	30
2.5.1 Candidate Gene Association Studies .....	30
2.5.1.1 Candidate Genes Associated with Chromosome Segregation .....	32
2.5.1.2 Genes Association with Human Genome-Wide Recombination Counts (Shown in the Bottom Half of Table 2.2, Below the Double Line) .....	40
2.5.2 Strongest Results from the Genome-Wide Association Study .....	43
2.5.2.1 rs10948101 on chromosome 6 near <i>VEGFA</i> .....	47
2.5.2.2 rs11535058 on chromosome 2 near <i>SLC39A10</i> .....	49
2.5.2.3 rs35288347 on chromosome 19 near <i>AURKC</i> .....	51
2.5.2.4 rs9984132 on chromosome 21 located in a gene rich region .....	53
2.5.2.5 rs73178888 on chromosome 8 near a region associated with meiotic recombination .....	55
2.5.2.6 rs115281615 on chromosome 4 near <i>CPEB2</i> .....	57

2.5.2.7 rs2560850 on chromosome 5 in an intron of <i>MYO10</i> .....	59
2.5.3 Strongest Results from the TDT Analyses .....	61
2.6 DISCUSSION.....	62
2.6.1 Genes Associated with Cohesin Complex .....	62
2.6.2 Genes Associated with the Synaptonemal Complex (SC).....	63
2.6.3 Association with Recombination-Related Variants .....	64
2.6.4 Gene Discovery .....	66
2.6.5 Conclusion and Future Directions.....	66
2.7 ACKNOWLEDGMENTS.....	67
<b>3.0 ANALYSES STRATIFIED BY MATERNAL AGE AND RECOMBINATION</b>	
<b>FURTHER CHARACTERIZE GENES ASSOCIATED WITH MATERNAL</b>	
<b>NONDISJUNCTION OF CHROMOSOME 21 .....</b>	<b>68</b>
3.1 CHAPTER OVERVIEW.....	68
3.2 INTRODUCTION .....	69
3.3 METHODS.....	71
3.3.1 Study Participants and Ethics Statement .....	71
3.3.1.1 Participants .....	71
3.3.1.2 Ethics Statement .....	71
3.3.2 Genotyping, Imputation, and Population Structure.....	72
3.3.3 Determining Meiotic Stage of Error and Recombination on Chromosome 21	
.....	73
3.3.4 Association Analyses .....	74
3.4 RESULTS.....	80



<b>3.4.1 Candidate Loci .....</b>	<b>83</b>
<b>3.4.1.1 Results for Candidate Loci Associated with Genome-Wide Recombination Counts .....</b>	<b>85</b>
<b>3.4.1.2 Results for Candidate Loci Involved with Meiotic Processes .....</b>	<b>87</b>
<b>3.4.2 Prior chromosome 21 nondisjunction GWAS associations.....</b>	<b>89</b>
<b>3.4.2.1 Results for Top Associations from Previous All Maternal (MI and MII) Errors Combined Analysis.....</b>	<b>92</b>
<b>3.4.2.2 Results for Top Associations from Previous Maternal MI Errors Analysis.....</b>	<b>93</b>
<b>3.4.2.3 Results for Top Associations from Previous Maternal MII Errors Analysis.....</b>	<b>95</b>
<b>3.4.2.4 Results for Top Associations from Previous Maternal MI Errors vs. MII Errors Analysis .....</b>	<b>96</b>
<b>3.5 DISCUSSION.....</b>	<b>97</b>
<b>3.6 CONCLUSION.....</b>	<b>102</b>
<b>3.7 APPENDIX: METHODS FOR CALLING MEIOTIC STAGE OF NONDISJUNCTION AND RECOMBINATION EVENTS .....</b>	<b>103</b>
<b>3.7.1 Overview .....</b>	<b>103</b>
<b>3.7.2 Calling Stage of Origin in Trios.....</b>	<b>104</b>
<b>3.7.3 Calling Recombination Events in Trios .....</b>	<b>105</b>
<b>3.7.4 Overview of Inference in Dyads.....</b>	<b>108</b>
<b>3.7.5 Calling Stage of Origin in Dyads .....</b>	<b>108</b>
<b>3.7.6 Calling Recombination Events in Dyads.....</b>	<b>109</b>

<b>3.8 ACKNOWLEDGMENTS.....</b>	<b>114</b>
<b>4.0 CONCLUSION .....</b>	<b>115</b>
<b>4.1 SUMMARY OF FINDINGS.....</b>	<b>115</b>
<b>4.2 STRENGTHS AND LIMITATIONS .....</b>	<b>116</b>
<b>4.3 FUTURE ANALYSES .....</b>	<b>116</b>
<b>BIBLIOGRAPHY .....</b>	<b>119</b>

## List of Tables

<b>Table 2.1 Description of primary analyses and sample sizes .....</b>	<b>21</b>
<b>Table 2.2 Candidate gene results.....</b>	<b>31</b>
<b>Table 2.3 Top hits from the all mothers vs. fathers genome-wide association study.....</b>	<b>43</b>
<b>Table 2.4 Top hits from the MI mothers vs. fathers genome-wide association study.....</b>	<b>44</b>
<b>Table 2.5 Top hits from the MII mothers vs. fathers genome-wide association study .....</b>	<b>45</b>
<b>Table 2.6 Top hits from the MI mothers vs. MII mothers genome-wide association study.</b>	<b>45</b>
<b>Table 2.7 Top hits from the TDT (transmission disequilibrium test) .....</b>	<b>46</b>
<b>Table 3.1 Definition of stratified analysis subgroups and associated sample sizes.....</b>	<b>75</b>
<b>Table 3.2 Summary of selected results.....</b>	<b>81</b>

## List of Figures

Figure 1.1 Meiotic error mechanisms .....	2
Figure 1.2 Inferring parent of origin in a trio .....	5
Figure 1.3 Determining meiotic stage of error in a trio.....	6
Figure 1.4 Examples of recombination profiles for two trios .....	7
Figure 1.5 Examples of recombination calls in two trios with genotyping “noise” .....	8
Figure 1.6 Informative markers in dyads .....	10
Figure 1.7 Informative markers on chromosome 21 in a dyad.....	11
Figure 1.8 Using a moving average to call recombination events in a dyad.....	12
Figure 2.1 LocusZoom plot for <i>RAD21L</i> .....	34
Figure 2.2 LocusZoom plot for <i>SYCE2</i> .....	36
Figure 2.3 LocusZoom plot for <i>SYCP1</i> .....	38
Figure 2.4 LocusZoom plot for <i>SYCP2</i> .....	39
Figure 2.5 LocusZoom plot for <i>SMEK1</i> .....	42
Figure 2.6 LocusZoom plot for <i>VEGFA</i> locus .....	48
Figure 2.7 LocusZoom plot for <i>SLC39A10</i> locus.....	50
Figure 2.8 LocusZoom plot for <i>AURKC</i> locus .....	52
Figure 2.9 LocusZoom plot for rs9984132 locus .....	54
Figure 2.10 LocusZoom plot for <i>DLGAP2</i> locus .....	56
Figure 2.11 LocusZoom plot for <i>CPEB2</i> locus .....	58
Figure 2.12 LocusZoom plot for <i>MYO10</i> locus.....	60
Figure 3.1 Distribution of maternal ages by meiotic stage of error .....	76

<b>Figure 3.2 Distributions of positions of single MI recombination events and most proximal MII events observed on chromosome 21.....</b>	<b>78</b>
<b>Figure 3.3 Summary of stratified analysis results for 37 candidate loci.....</b>	<b>84</b>
<b>Figure 3.4 LocusZoom plots for selected associations in candidate loci for global recombination (<i>CCDC43</i>, <i>RAD21L</i>, and <i>CPLX1/RNF212</i> loci) .....</b>	<b>86</b>
<b>Figure 3.5 LocusZoom plots for selected associations in candidate genes involved in meiotic processes (<i>SYCP1</i>, <i>SYCP2</i>, and <i>MND1</i>).....</b>	<b>88</b>
<b>Figure 3.6 Summary of stratified analysis results for 48 previous nondisjunction GWAS loci .....</b>	<b>90</b>
<b>Figure 3.7 Forest plots for four selected prior GWAS loci .....</b>	<b>91</b>
<b>Figure 3.8 Overview of methods for calling recombination in trios and dyads .....</b>	<b>104</b>
<b>Figure 3.9 Method for calling recombination events in a full trio.....</b>	<b>106</b>
<b>Figure 3.10 Scoring recombination calls in the training set trios.....</b>	<b>110</b>
<b>Figure 3.11 Method for calling recombination events in a masked-trio dyad .....</b>	<b>112</b>
<b>Figure 3.12 Optimizing recombination calls on the pseudo-dyads.....</b>	<b>113</b>

## **1.0 INTRODUCTION**

### **1.1 BACKGROUND AND MOTIVATION**

#### **1.1.1 Meiosis, Aneuploidy, and Established Risk Factors**

Human reproduction relies on the generation of haploid gametes during meiosis in order to form euploid offspring. But, due to meiotic nondisjunction, at least 10% (and potentially up to a quarter) of human conceptions are aneuploid, with the majority not surviving to term [1-5].

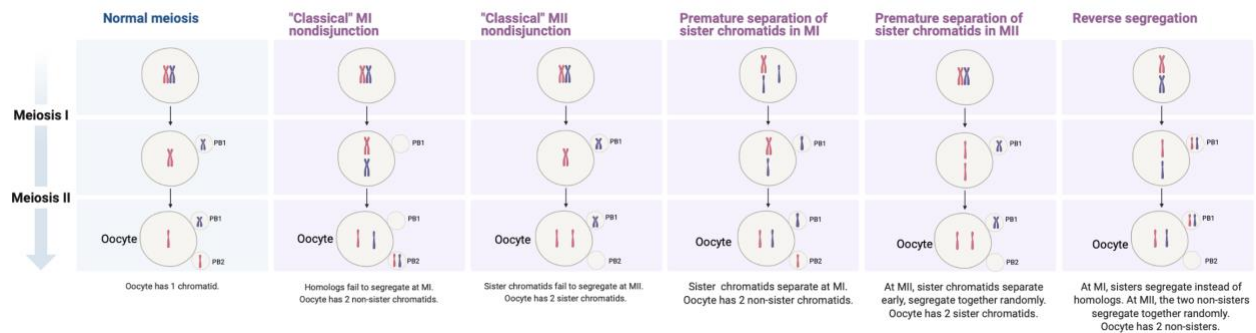
The higher rate of nondisjunction in women than in men reflects the different timescales on which meiosis operates in men and women. Oogenesis begins during development and pauses during prophase of meiosis I (MI) after recombination, with resumption at ovulation and completion of meiosis II (MII) after fertilization. Hence the duration of an oocyte's arrest is essentially equal to a woman's age at conception, and the risk of a meiotic error increases exponentially with age. Spermatogenesis, in contrast, begins at puberty, without a long period of arrest. (This topic is reviewed in [6].)

Trisomy 21 (Down syndrome) is a useful condition to study for gaining insight into meiotic nondisjunction, since most other aneuploidies are not compatible with survival to term. Among live-born children with trisomy 21, over 90% of the nondisjunction events are maternal in origin, with about 75% due to apparent MI errors and 25% due to apparent MII errors [7]. Given the different timescales involved with completion of maternal MI and MII, it's unsurprising that MI and MII nondisjunction have different risk factors.

In particular, altered patterns of recombination are a well-established risk-factor for maternal nondisjunction of most chromosomes, including chromosome 21 [8]. Specific to maternal chromosome 21 nondisjunction, both MI and MII errors are associated with altered recombination [9-13]. About half of MI errors show no recombination on chromosome 21, and those with a single exchange usually show telomeric recombination [9, 11, 14]. MII errors are associated with pericentromeric recombination [9, 11, 13, 14].

### 1.1.2 Possible Mechanisms and Genes of Interest

As meiosis is a complex process with multiple steps and critical proteins, there are several opportunities for errors to occur that can result in trisomy. Here we review several mechanisms (see Figure 1.1, which was adapted from “Non-disjunction in Meiosis” by BioRender.com (2020) and retrieved from <https://app/biorender.com/biorender-templates>).



**Figure 1.1 Meiotic error mechanisms**

Conceptually, the simplest errors are “classical” MI or MII errors (often referred to in this dissertation and elsewhere simply as MI and MII nondisjunction/errors). In a classical MI error, a pair of homologs fails to disjoin at MI. Then, at MII, each homolog splits into two sister chromatids, so that the oocyte ends up with two non-sister chromatids. In a classical MII error,

homolog-pairs separate normally at MI, but sister chromatids fail to separate at MII, resulting in an oocyte with two sister chromatids.

Another mechanism is premature/precocious separation of sister chromatids (PSSC). When sister chromatids separate too soon, they may segregate at random, independently of each other instead of to opposite poles in a coordinated fashion. If this occurs during MI, the oocyte may contain two non-sister chromatids, mimicking a classical MI error. If it occurs during MII, the oocyte may contain two sister chromatids, mimicking a classical MII error.

Finally, reverse segregation (RS) occurs when sister chromatids segregate at MI instead of homologs, leading to a euploid intermediate containing two non-sister chromatids. At MII these may move together to the oocyte, resulting in an aneuploidy mimicking a classical MI error.

Note that although PSSC and RS may mimic classical MI and MII errors, all of these mechanisms are distinct: chromatids or chromosomes fail to separate vs. separate too soon. The technique used in this dissertation cannot distinguish all of these error types from each other, but only whether there was an “apparent MI error” or an “apparent MII” error, according to whether the other passed two different or two identical copies of chromosome 21 to the child. Therefore, while distinct genetic risk factors may influence each mechanism, this study is not designed to prove or disprove which mechanism or variants caused each instance of aneuploidy. Patterns of association observed across different subgroups and the functions of implicated genes may nevertheless provide insight into these questions.

To complement the genome-wide scan we performed for maternal meiotic nondisjunction, we also tested a set of candidate loci. These comprise well-established meiosis protein genes (e.g., synaptonemal complex and cohesin subunits) as well as loci associated with genome-wide



recombination counts in an Icelandic population [15] (see [2.3](#) for a discussion of the candidate loci).

## **1.2 GENOTYPING ARRAY DATA IN THE CONTEXT OF DOWN SYNDROME**

In order to find genetic variants affecting the risk of maternal nondisjunction of chromosome 21 and its associated recombination patterns, we studied a group of children with Down syndrome and their parents. Genotype array data from these families serves several purposes in this study. First, chromosome 21 genotypes enable us to determine which parent contributed the child's extra copy of chromosome 21 and the stage of meiosis in which the error apparently occurred (MI vs. MII). Second, we can infer the locations of recombination events on chromosome 21, with the goal of categorizing nondisjunction events by recombination profile (discussed above as a known risk factor). Finally, genome-wide data allow us to systematically test the genome (and candidate regions) for association with nondisjunction, while stratifying by stage of error and recombination profile.

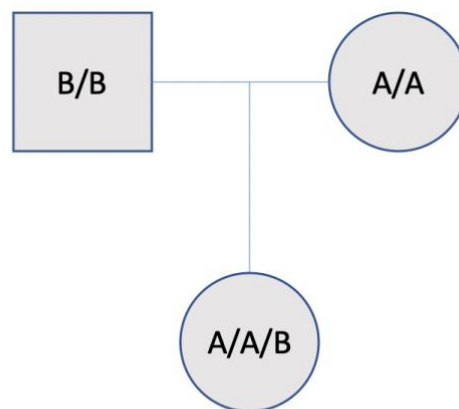
### **1.2.1 Family Data and Informative Markers**

The parent and (with a few assumptions) the meiotic stage of origin for each nondisjunction event can be inferred by comparing the parents' and children's genotypes at markers along chromosome 21. For trios, the parents' and children's genotypes are all known, and the procedure is a straightforward extension of methods that have been used in previous studies using sparser

marker sets (see, e.g., [7, 16]). For mother-child dyads, which contain less information, a new method is described below.

### 1.2.1.1 Methods for Trios

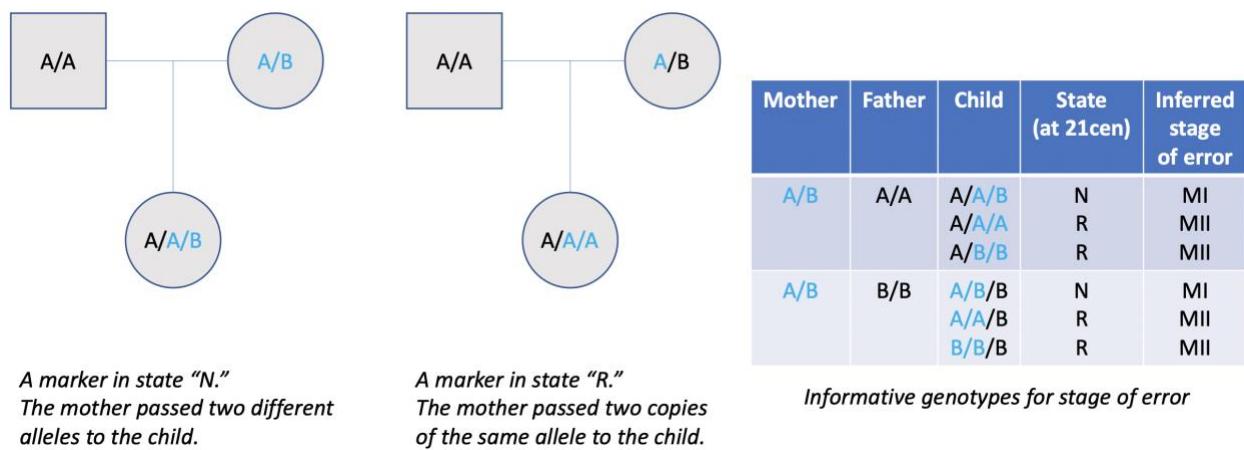
To determine the parent of origin in a trio, we compare the child's genotypes on chromosome 21 to the parents'. A marker (here, usually a SNP) is informative when one parent has genotype A/A and the other has genotype B/B (i.e., the parents are homozygous with different alleles; see Figure 1.2). In that case, the nondisjoining parent must transmit two identical alleles to the child, and the other parent transmits one copy of the other allele. Hence if the child's genotype is A/A/B, then the nondisjoining parent is expected to be the parent with genotype A/A. Since any particular genotype is uncertain, here we used all informative SNPs on chromosome 21 and considered the proportion consistent with maternal nondisjunction. Allowing for genotyping error and mosaicism, we can determine that the mother was the nondisjoining parent whenever the proportion is near 1.



*Infer that the mother is the parent of origin, since she must have passed A/A to the child.*

**Figure 1.2 Inferring parent of origin in a trio**

Next, similar logic is used to infer the meiotic stage of origin for the trisomy in each trio. In MII nondisjunction, sister chromatids nondisjoin, so that the mother transmits two identical alleles to the child. In MI nondisjunction, homologous chromosomes nondisjoin, so that non-identical alleles are transmitted. Hence a SNP is informative for the stage of error if the genotypes enable us to determine whether the two alleles passed from the mother to the child are identical or not. This occurs when the mother is heterozygous (A/B) and the other parent the father is homozygous (A/A, say). Figure 1.3 summarizes inference for stage of error. When the mother has passed two identical alleles for a SNP, the SNP is scored as reduced to homozygosity (R). If the two alleles are different, it is score as nonreduced (N).

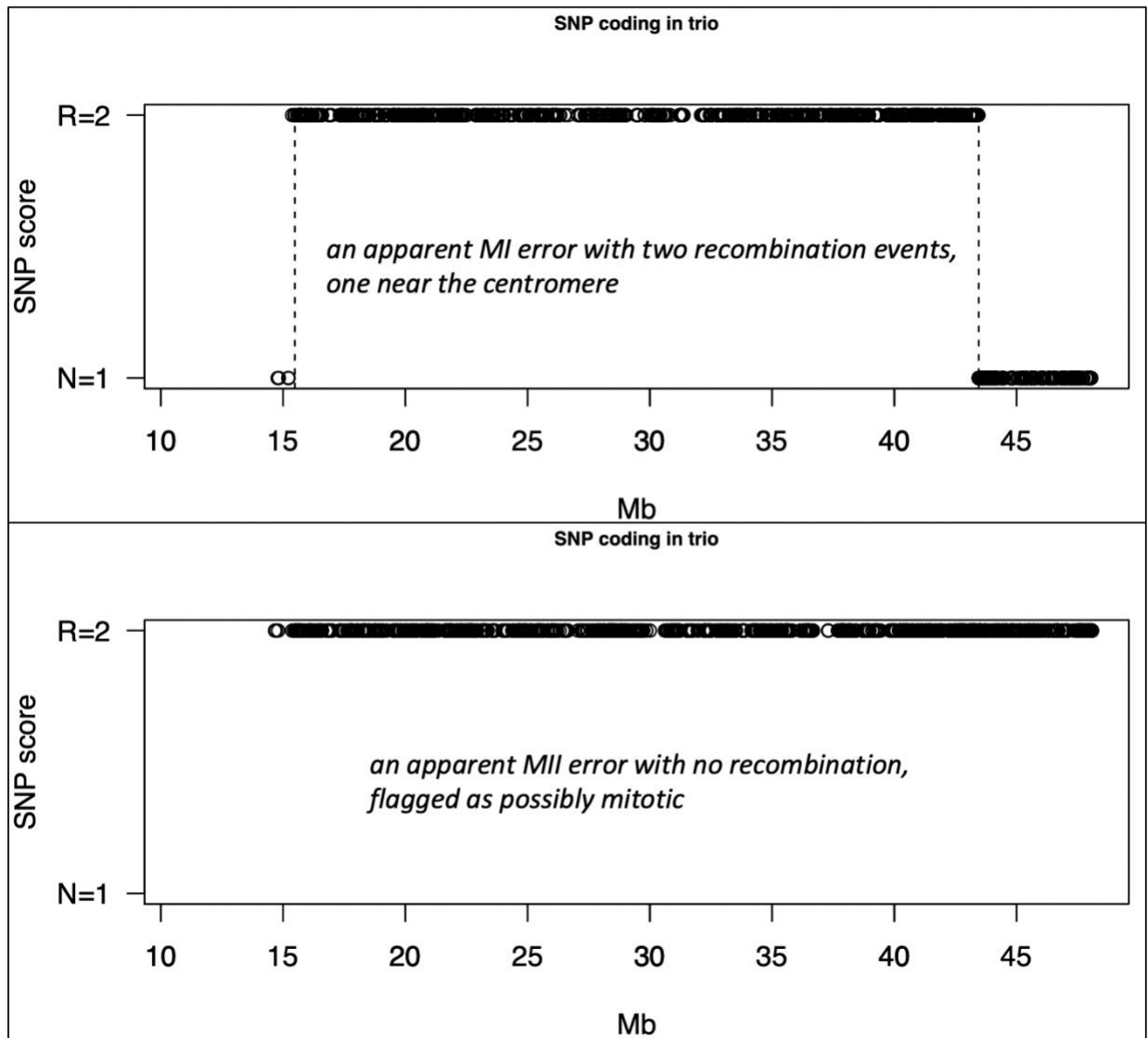


**Figure 1.3 Determining meiotic stage of error in a trio**

In addition to assuming implicitly a meiotic (rather than mitotic) origin for each trisomy, we also assume that no recombination events are close enough to the centromere to go unobserved (i.e., more proximal than the nearest informative SNP). Then the zygosity (R or N) of the informative 21q SNPs nearest the centromere reflects the stage of error, R indicating an MII error and N indicating an MI error (21p SNPs are generally not informative).

Plotting the state of each informative SNP (1 for N, 2 for R) vs. its physical chromosomal position, recombination events appear as changes in the “level” of the graph (e.g., ...11112222...),

as in the upper panel of Figure 1.4. Manually inspecting the plots, we can flag apparent MII cases without recombination (i.e., graphs with  $y = 2$  for all SNPs, except for a few genotyping errors); these are often considered potentially mitotic and therefore excluded from analysis (Figure 1.4, lower panel).



**Figure 1.4** Examples of recombination profiles for two trios

Calling recombination events involves a tradeoff between over- and under-sensitivity. A “switch” from R to N (or vice-versa) suggests a recombination event, but not every switch really signifies an event (Figure 1.5). On the one hand, a pattern like ...NNNNNNRNNNN... would

indicate two recombination events very close to each other (“tight double recombination”), which is unlikely, given the relative likelihood of genotyping error (either in the child or one of the parents) and that of a true tight double recombination (given crossover interference). On the other hand, in a pattern like ...NNNNNRRRRRNNNNN..., technical error is less likely to account for the apparent double recombination. More SNPs in the “middle segment” make the double recombination more plausible.

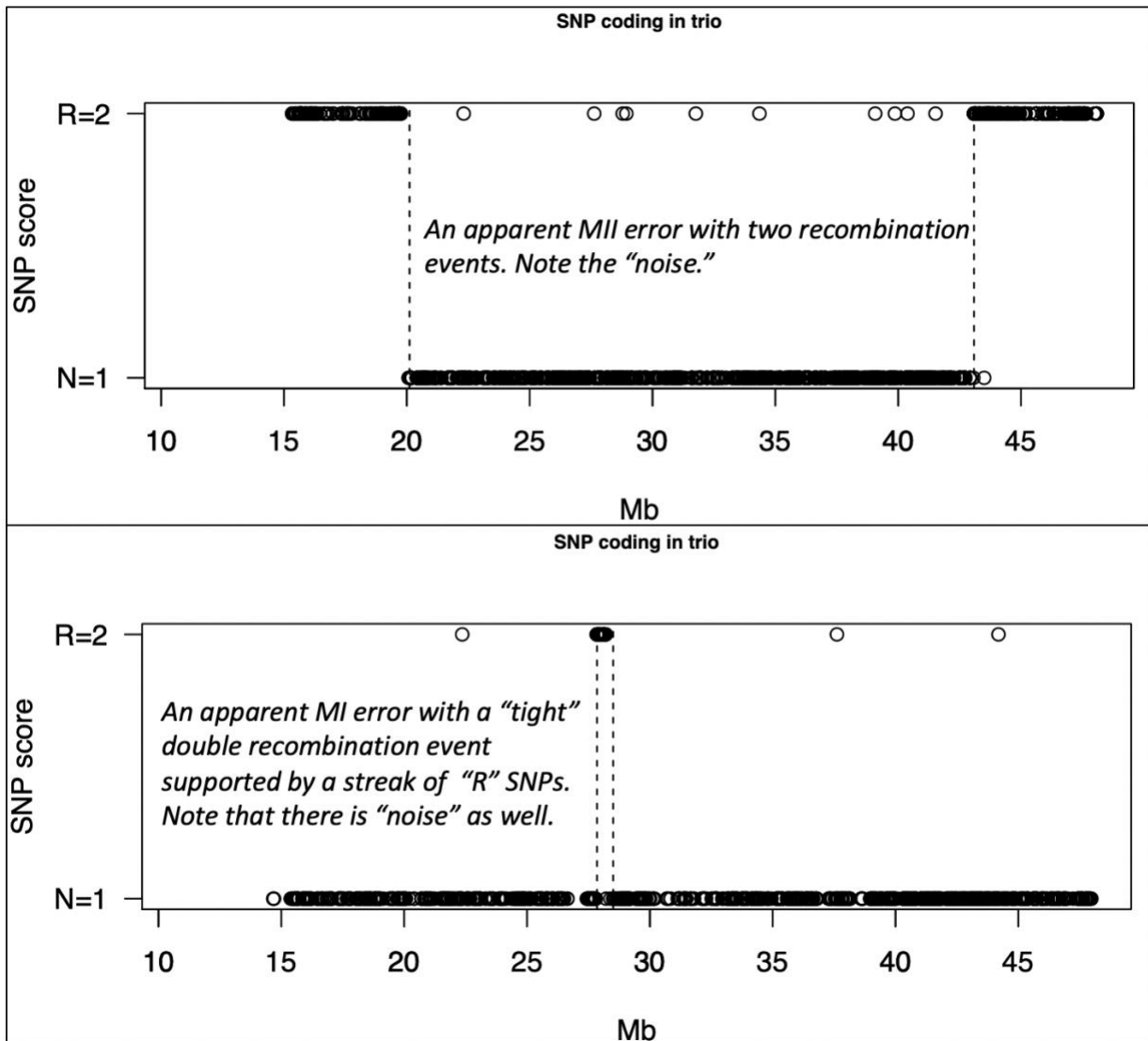


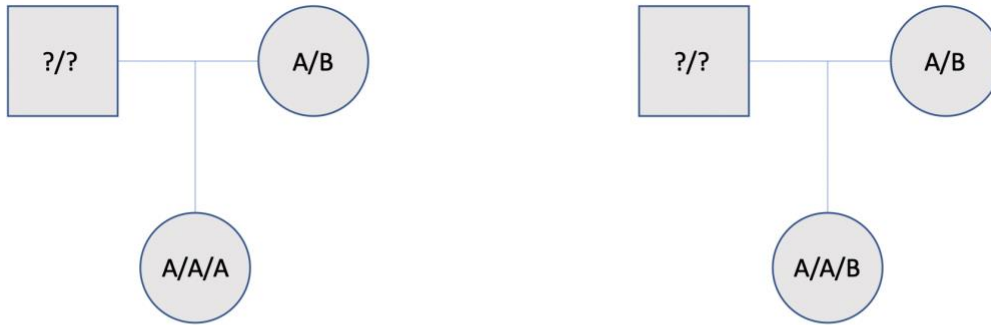
Figure 1.5 Examples of recombination calls in two trios with genotyping “noise”

### 1.2.1.2 Methods for Dyads

The approach described above for trios requires the genotypes of both parents, in addition to the child's. Inference is still possible in a genotyped parent-child dyad, where one parent's genotypes are unknown. Here we specifically consider mother-child dyads, in which the father's genotypes are unknown (the mother's nondisjunction and recombination phenotypes are analyzed downstream).

As in trios, the parent of origin is first confirmed to be the mother in each case by considering the proportion of SNPs consistent with maternal nondisjunction.

As in trios, a SNP is informative for stage of error only if the mother is heterozygous (A/B). If the child is homozygous (A/A/A or B/B/B), we infer that the mother transmitted two identical alleles and score the SNP as R. If the child's genotype is A/A/B or A/B/B, the SNP is partially informative, since it can be inferred that the mother did not transmit B/B or A/A (respectively). In that case, the SNP is scored as X. Figure 1.6 summarizes inference in dyads about meiotic stage of origin.



A marker in state “R” in a dyad.  
The mother must have passed A/A to the child.

Example of an “X” marker in a dyad.  
The mother may have passed A/A or A/B, but not B/B.

Mother	Father	Child	State (at 21cen)	Inferred stage of error
A/B	Unknown	A/A/A	R	MII
		A/A/B	X (unknown)	MI (if X:R ratio is high near 21cen)
		A/B/B	X (unknown)	MI (if X:R ratio is high near 21cen)
		B/B/B	R	MII

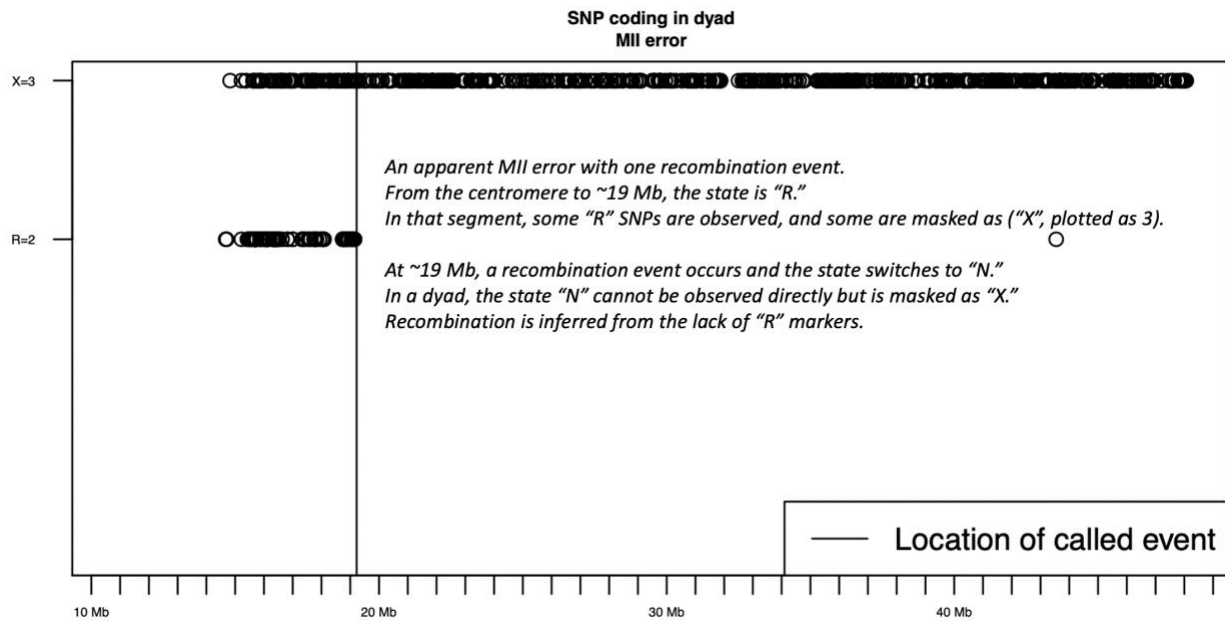
*Informative genotypes for stage of error in dyads*

**Figure 1.6 Informative markers in dyads**

For each dyad, we obtain a string of Rs and Xs representing the states of informative (or partially informative) SNPs near the centromere. As in trios, a string of Rs near the centromere is strong evidence of an MII error. But since SNPs in the N state can’t be observed directly (they’re masked as Xs, along with some R SNPs), we can infer MI errors only from the absence of R SNPs. Hence a higher ratio of Rs to Xs in a string is stronger evidence of an MII error (and a lower ratio is stronger evidence of an MI error).

Similarly, recombination events are reflected by changes in the density of R vs. X along the chromosome. As in the case for trios, there is a problem of over- vs. under-calling recombination events, which is further complicated by the lower density of information. Although no particular X can be “unmasked” with certainty, the density of Rs nearby provides information. For instance, in the pattern ...RRXR..., the X is likely to mask an R (an N would imply tight double recombination). But in a pattern such as ...RRRRRXXXXXXXXXX..., the true state of

some of the Xs is probably N (if the state were truly R at each position, it would be unlikely to observe a long sequence of Xs). Hence a recombination event likely occurred. See Figure 1.7.



**Figure 1.7 Informative markers on chromosome 21 in a dyad**

Using this intuition, a higher density of Xs is evidence that the true state is N; recombination can be inferred when the density of Xs changes enough (by some measure). In this work we propose using a moving average to identify recombination events, calculating for each SNP the proportion of nearby informative SNPs scored X (Figure 1.8). This method leaves the number of SNPs to average over and the threshold as parameters to be optimized. To select the optimal parameters for this method, we apply it to the full-data trios, treating the fathers' genotypes as missing and the recombination calls in the trios as correct. Full details are given in [Section 3.7](#).



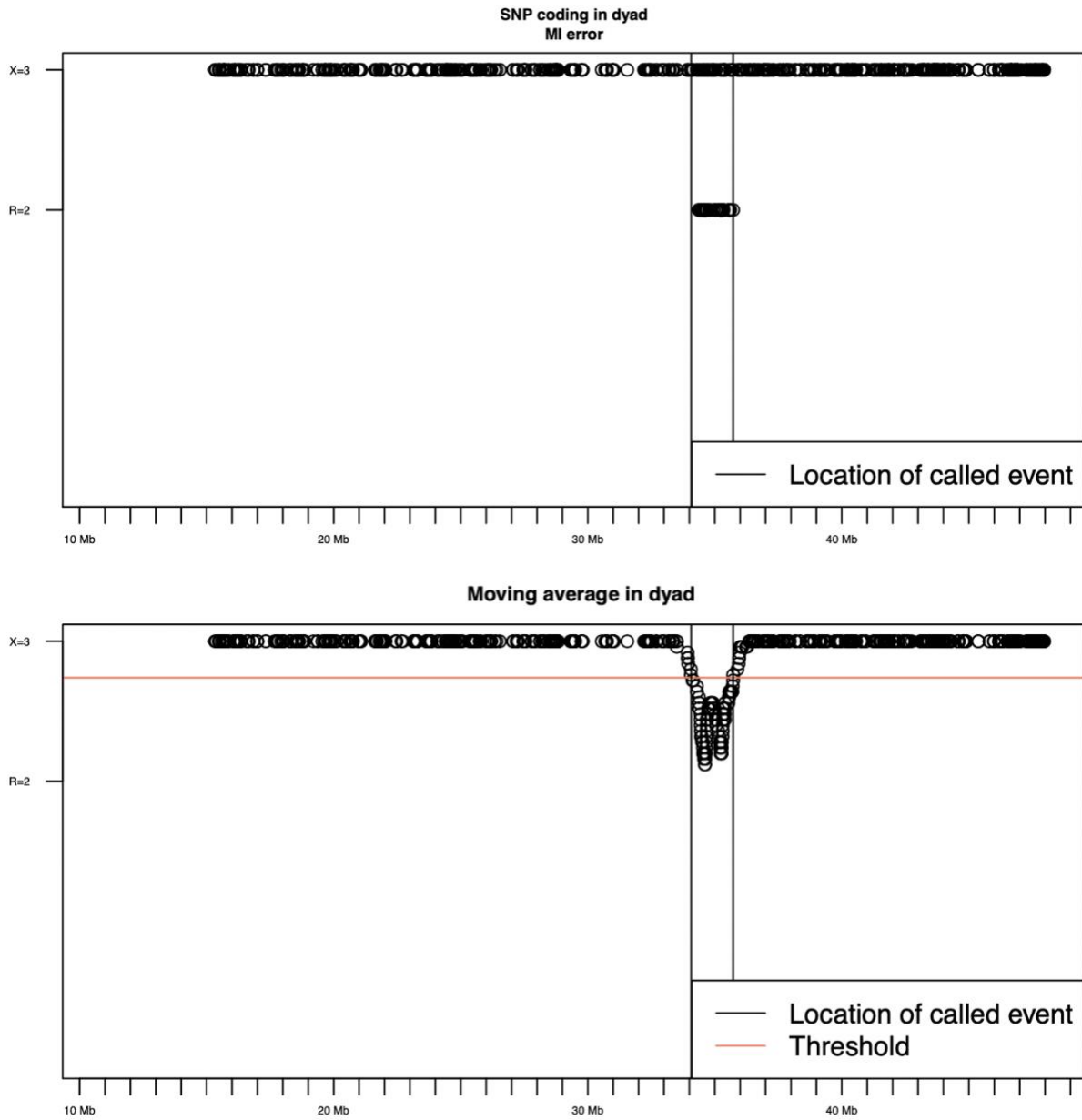


Figure 1.8 Using a moving average to call recombination events in a dyad

## 1.2.2 Description of Data Set

This study includes 2,186 subjects, comprising 749 children with trisomy 21 and most of their parents. Subjects were initially recruited using birth surveillance through the National Down Syndrome Project [7], a multi-site, population-based study, and later using convenience sampling.

Trisomy was confirmed by karyotyping; both full and mosaic trisomies are included, but not translocation trisomies. Genotyping was performed with the Illumina HumanOmniExpressExome-8v1-2 array at CIDR, and after quality assurance/control about 1 million single nucleotide polymorphism (SNP) genotypes were available. Trisomic genotypes for SNPs on chromosome 21 were called with previously-developed methods [17]. After imputation using the 1000 Genomes Project's Phase I data and subsequent filtering, genotypes were available for about 9 million SNPs. Short tandem repeat (STR) marker genotypes were also available in some cases; STR genotyping was done historically as this cohort was recruited primarily to determine the parent and stage of origin of the trisomy.

The dataset includes chromosome 21 SNP genotypes for 630 parent-child trios, 95 mother-child dyads, and 17 father-child dyads.

## **1.3 AIMS AND SUMMARY OF THIS DISSERTATION**

### **1.3.1 Finding Genetic Risk Factors for Nondisjunction**

The first aim of this dissertation is to identify genetic risk factors for maternal meiotic nondisjunction of chromosome 21 in either stage of meiosis. [Chapter 2](#), which consists of a published genome-wide association and candidate gene study (see [18]), accomplishes this. Briefly, we performed a case-control study in which the mothers of children with Down syndrome are treated as cases and the fathers as controls. We found evidence that common variants at loci involved with oocyte maturation and meiotic processes may influence nondisjunction risk, with some specific to MI or MII.

### **1.3.2 Using Family Data to Call Recombination Events on Chromosome 21**

The second aim is to develop and implement methods for calling recombination events on chromosome 21 in parent-child trios and dyads. For trios this is quite straightforward; for mother-child dyads where the fathers' genotypes are unavailable, it is more challenging. As there is no "gold standard" for either case, we implement a simple method for trios and then use the resulting recombination calls as a "training set" to tune parameters in a novel method for calling recombination the dyads. We attain a high accuracy on the training set and conclude that it is feasible to call recombination events in mother-child dyads, resulting in a larger usable sample for association analyses (and potentially lower cost in future studies). Details of these recombination-calling methods are given in the [Appendix of Chapter 3](#), which has been published as part of [19].

### **1.3.3 Characterizing Genetic Risk Factors for Nondisjunction**

Our third aim is to dissect the associations discovered in the first aim. As apparent MI and MII errors may occur through different mechanisms, there may be genetic heterogeneity within each (apparent) type of error. Therefore we used the recombination calls from the second aim to classify each maternal nondisjunction event by its recombination profile on chromosome 21. This enabled us to perform stratified analyses of the loci found in the first aim, using the most important nondisjunction risk factors as stratification variables: maternal age and recombination profile. We find evidence consistent with genetic heterogeneity and attempt to interpret our results in the context of possible mechanisms. The third aim is accomplished in [Chapter 3](#), consisting of a second published study (see [19]).

## 2.0 A CANDIDATE GENE ANALYSIS AND GWAS FOR GENES ASSOCIATED WITH MATERNAL NONDISJUNCTION OF CHROMOSOME 21

This chapter has been published in PLoS Genetics. Per PLoS license and copyright policy (<https://journals.plos.org/plosgenetics/s/licenses-and-copyright>), the article is reproduced here under a Creative Commons BY license with minor formatting and non-scientific changes. The original article and supporting information are available online at <https://doi.org/10.1371/journal.pgen.1008414>.

Jonathan M. Chernus<sup>1</sup>, Emily G. Allen<sup>2</sup>, Zhen Zeng<sup>3</sup>, Eva R. Hoffman<sup>4</sup>, Terry J. Hassold<sup>5</sup>, Eleanor Feingold<sup>1,3\*</sup>, Stephanie L. Sherman<sup>2\*</sup>

<sup>1</sup> Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. <sup>2</sup> Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America. <sup>3</sup> Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. <sup>4</sup> Center for Chromosome Stability, University of Copenhagen, Copenhagen, Denmark. <sup>5</sup> School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, Washington, United States of America. \* EF and SLS are joint senior authors.

## **2.1 CHAPTER OVERVIEW**

Human nondisjunction errors in oocytes are the leading cause of pregnancy loss, and for pregnancies that continue to term, the leading cause of intellectual disabilities and birth defects. For the first time, we have conducted a candidate gene and genome-wide association study to identify genes associated with maternal nondisjunction of chromosome 21 as a first step to understand predisposing factors. A total of 2,186 study participants were genotyped on the HumanOmniExpressExome-8v1-2 array. These participants included 749 live birth offspring with standard trisomy 21 and 1,437 parents. Genotypes from the parents and child were then used to identify mothers with nondisjunction errors derived in the oocyte and to establish the type of error (meiosis I or meiosis II). We performed a unique set of subgroup comparisons designed to leverage our previous work suggesting that the etiologies of meiosis I and meiosis II nondisjunction differ for trisomy 21. For the candidate gene analysis, we selected genes associated with chromosome dynamics early in meiosis and genes associated with human global recombination counts. Several candidate genes showed strong associations with maternal nondisjunction of chromosome 21, demonstrating that genetic variants associated with normal variation in meiotic processes can be risk factors for nondisjunction. The genome-wide analysis also suggested several new potentially associated loci, although follow-up studies using independent samples are required.

## **2.2 AUTHOR SUMMARY**

Approximately one of every 700 babies is born with trisomy 21 - an extra copy of chromosome 21. Trisomy 21 is caused by the failure of chromosomes to segregate properly during

meiosis, generally in the mother. Past studies have defined altered patterns of recombination along nondisjoined chromosomes as risk factors for human nondisjunction and model systems have clearly shown that specific genes involved recombination and other early meiotic processes play a role in the fidelity of chromosome segregation. However, no genome-wide genetic study (GWAS) has ever been conducted using maternal human nondisjunction as the disease phenotype. This study takes the first step to understand predisposing factors. We used chromosome 21 genotypes from the parents and child to identify mothers with nondisjunction errors derived in the oocyte and to establish the type of error (meiosis I or meiosis II). We then conducted a unique set of subgroup comparisons designed to leverage our previous work that shows that the etiologies of meiosis I and meiosis II nondisjunction differ for trisomy 21. Both the candidate gene study and the GWAS provide evidence that meiotic-specific structures and processes are vulnerable to genetic variants that lead to increased risk of human chromosome nondisjunction.

## **2.3 INTRODUCTION**

Correct segregation of chromosomes during the two successive meiotic divisions is essential for the formation of haploid gametes. At least 10% of human pregnancies produce aneuploid embryos with too many or too few chromosomes, the majority of which are lost during pregnancy. If they survive to term, many have severe congenital defects and developmental and intellectual disability. Thus, meiotic nondisjunction is the leading cause of pregnancy loss and birth defects in humans and an important limiting factor in women's reproductive life span. (reviewed in [1, 2, 4, 5, 13]).

In humans, meiosis in females is highly prone to chromosome segregation errors i.e., nondisjunction or premature separation of sister chromatids (PSSC) and these errors increase exponentially with increasing maternal age. The differences between the development of oocytes and sperm clearly influence susceptibility for meiotic nondisjunction. Most importantly, they work on different timelines. In both sexes, meiosis starts with an initial step of DNA replication and the establishment of sister chromatid cohesion, followed by synapsis and recombination between homologous chromosomes. Homologs then separate at the end of meiosis I (MI), whereas sister chromatids separate in meiosis II (MII). Spermatogenesis begins after puberty and cells entering meiosis move from one stage to the other without delay. In contrast, oogenesis begins during fetal development and is arrested in prophase I after chromosomes synapse and recombine. MI resumes in the woman's adult life just before the ovulation; MI is completed and the first polar body is extruded. MII begins but arrests for a short period as the oocyte travels through the Fallopian tubes, and is only completed if the oocyte is fertilized. Thus, meiosis in females extends over a 10 to 50 year period; the age of the woman at conception reflects the age of the oocyte, and basically the period of arrest in MI. Given the mechanistic differences and temporal separation of maternal MI and MII, it is not surprising that associated risk factors differ for MI and MII nondisjunction errors (reviewed in [6]).

Trisomy 21 has become an important resource to understand meiotic nondisjunction in humans, as it is one of the few aneuploid conditions that survives to term. However, even for trisomy 21, involving the smallest human autosome, about 50-80% conceptions are estimated to be lost during pregnancy [5, 20]. Using chromosome 21 genetic markers to categorize the type of meiotic error among live births with trisomy 21, over 90% are derived from errors in the oocyte, of which at least 75% are estimated to be initiated in MI and about 25% in MII (e.g., [7]).

In this study, our goal was to discover genetic variants that increase the risk for maternal nondisjunction of chromosome 21 using both a candidate gene approach and a genome-wide association study. We focused on candidate genes that have been associated with chromosome dynamics early in meiosis. Accurate segregation depends on the coordinated control of sister chromatid cohesion with chromosome synapsis and the assembly of the synaptonemal complex (SC) and, within these structures, meiotic recombination [21, 22]. Below we provide a brief overview of the role of some of the important meiotic genes that mediate these processes, and that we have examined in the present study. Bolcun-Filas and Schimenti [21] have summarized the meiotic defects that are observed in the associated mutant mouse models.

In a meiotic cell, DNA is organized as an array of loops along a proteinaceous axis. The axes are composed of the meiosis-specific synaptonemal complex, in association with condensin/cohesin complexes. Several of the components of meiotic cohesin are meiosis-specific, including those encoded by *SMC1 $\beta$* , *REC8*, *RAD21L*, and *STAG3*. The SC brings homologous chromosomes into close proximity and promotes recombination and chiasmata formation [23]. The mature SC is a tripartite structure, composed of two parallel axial/lateral elements that bind to each homolog and a central element, with transverse filaments joining the individual axial/lateral elements [24, 25]. SYCP2 and SYCP3 are components of the axial/lateral elements. SYCP1 is a component of the transverse filaments and components of the central element are encoded by *SYCE1*, *SYCE2*, *SYCE3*, and *TEX12*. In addition to these structural sub-units, *HORMAD1* and *HORMAD2* code for proteins that load onto axes of meiotic chromosomes throughout early prophase I but are removed upon synapsis, a process that depends on the presence of TRIP13 [26]. In general, *HORMAD1* and *HORMAD2* play a role in coordinating progression of chromosome synapsis with meiotic recombination [27].



Meiotic recombination is initiated by programmed DNA double-strand breaks (DSBs) that occur as the meiotic chromosome axes develop early in prophase I. These breaks are generated by the SPO11 protein and its interacting partners MEI1, MEI4 and REC114 (reviewed in Cole et al. [28]). The DSBs are processed to generate single-stranded DNA that is bound by strand-exchange proteins DMC1 (meiosis specific) and RAD51 (ubiquitously expressed). The single-stranded DNA then engages in homology search. Proper function of DMC1 requires interactions with several meiotic accessory proteins, one of which is MND1. MND1, complexed with HOP2, stabilizes the DMC1 filaments on the resected end of the DSBs. This complex also increases the ability of the pre-synaptic filament to capture the double-stranded DNA (reviewed in Sansam and Pezza [29]).

As homologs synapse, so-called early recombination nodules transiently associate with ZMM proteins, including DNA mismatch repair proteins MSH4 and MSH5. Subsequently, a proportion of these are converted into late recombination nodules, detected by the mismatch repair proteins MLH1 and MLH3, and representing the vast majority of crossovers [30-34]. In addition, EXO1 and BLM function in crossover regulation, and with MLH1 and MLH3, appear to play a role in the crossover pathway that is subject to crossover interference (reviewed in Manhart and Alani [35]).

In addition to these candidate genes, we chose another group of genes that have been associated with the amount of global meiotic recombination in humans. The motivation for these candidates is based on the altered recombination patterns observed along nondisjoined chromosomes, a well-established predisposing factor for maternal nondisjunction of almost all human chromosomes studied to date (reviewed in [8]). Specifically for maternal chromosome 21 nondisjunction, altered meiotic recombination patterns are associated with both MI and MII error types [9-13]. For maternal MI-derived trisomy 21, about 40-47% of MI cases are derived from

oocytes with no meiotic exchange [9, 11, 14]. Among those with a single exchange, the majority of exchanges occur in the telomeric region of chromosome 21. MII errors are associated with pericentromeric exchanges [9, 11, 13, 14]. This apparent effect of an MI process – recombination – on MII nondisjunction suggests that at least a portion of so-called MII errors may have their origin in MI. In addition, there is evidence that genome-wide recombination counts in oocytes with a MI nondisjunction error of chromosome 21 are reduced compared to oocytes with normal meiosis [36, 37]. Also, previous studies indicate that oocyte-specific dysregulation of global recombination may contribute to the nondisjunction event [36]. Thus, we chose candidate genes identified in the largest GWAS study of meiotic recombination conducted on humans, a study based on 71,929 parent-offspring pairs from Iceland [15]. They found evidence for 13 variants in eight regions that were associated with genome-wide recombination counts.

For both the candidate gene and genome-wide association studies, we took a unique approach by using several different GWAS group comparisons (Table 2.1). These comparisons were crafted to address the likelihood that there are both distinct genetic factors influencing MI and MII nondisjunction and common factors affecting both. In addition, some of our analyses target the conflated phenotype of nondisjunction with survival to term. Study design issues are discussed in more detail below.

**Table 2.1 Description of primary analyses and sample sizes**

<b>Analysis</b>	<b>Analysis groups</b>	<b>Sample size</b>	<b>Contrast able to detect:</b>
Logistic regression	Mothers vs. fathers	705 vs. 645	Maternal NDJ and survival to term
...	MI mothers vs. fathers	535 vs. 645	MI-specific effects and survival to term
...	MII mothers vs. fathers	157 vs. 645	MII-specific effects and survival to term
...	MI mothers vs. MII mothers	535 vs. 157	MI- or MII-specific effects
TDT	All complete case trios	615 trios	Survival to term

## **2.4 METHODS**

### **2.4.1 Study Sample**

Our study participants included 749 live born offspring with free (non-translocation), maternally-derived trisomy 21 (both full and mosaic trisomy 21 were included) and their available biological parents. In almost all instances, the trisomy was confirmed by karyotype, although in some it was confirmed by birth record or parent report. Recruitment occurred in the U.S. by multiple sites since 1989, when the first population-based study was initiated. Recruitment for these population-based studies used birth surveillance systems to identify infants born with Down syndrome (details are provided in Freeman et al. [7]). Later, our recruitment strategy was not population-based, but instead a convenience sample of families with Down syndrome identified through our network of assessment sites, social and website media, and parent groups. Using self-reported race/ethnicity, 72% reported as White, 4% as Hispanic descent, 2% as African/African-American or Asian descent and about 23% with other or unknown descent.

#### **2.4.1.1 Ethics Statement**

Participants were recruited from several geographic areas with the collaboration of several institutions, including Arkansas (University of Arkansas for Medical Sciences, Arkansas Center for Birth Defects Research and Prevention, Arkansas Children's Hospital, Arkansas Reproductive Health Monitoring Systems), California (California Birth Defects Monitoring Program, Public Health Institute), Georgia (Department of Human Genetics, Emory University; Centers for Disease Control and Prevention), Iowa (University of Iowa, Registry for Congenital and Inherited Disorders), New Jersey (New Jersey Department of Health and Senior Services; Special Child

Health Services Registry; Eagleton Institute), and New York (New York State Department of Health Congenital Malformations Registry). Each recruitment site obtained IRB approval for their protocol, consent forms, and data sharing during the project period from their respective institutions. All samples were collected under written consent by each participant or their legal guardian. Emory University was the site for the data and biorepository. They obtained IRB approval for all sample processing and de-identified sample submission to the Center for Inherited Disease Research genotyping service (Emory School of Medicine IRB number IRB00005100). IRB approvals for genotyping samples and uploading to dbGaP were approved prior to the initiation of that genotyping project (dbGaP: phs000718).

#### **2.4.1.2 Genotyping**

DNA samples were obtained from lymphoblastoid cell lines (LCLs) (36.8%), saliva (23.7%), buffy coat (15.7%), whole blood (13%), unknown source (i.e., no record available) (8.4%), and buccal cell (0.09%). The remaining 2.2% of genotyped samples were HapMap controls derived from LCLs that were used by the Center for Inherited Disease Research (CIDR) for quality control (QC). The samples were genotyped in batches corresponding to 96-well plates and each plate contained two study duplicates and HapMap controls. Duplicates were randomly selected from all samples with sufficient DNA. Families were randomly distributed across plates with all members of each family on the same plate.

Genotyping was performed on the Illumina HumanOmniExpressExome-8v1-2 array by the Center for Inherited Disease Research (CIDR). The algorithm used for calling genotypes was GenomeStudio version 2011.1, Genotyping Module version 1.9.4 and GenTrain version 1.0. Genotype data that passed initial QC at CIDR were released to the Quality Assurance (QA)/QC analysis team at the University of Washington Genetics Coordinating Center (UWGCC) for data

cleaning and imputation. Details of these procedures can be found in Laurie et al. [38] and all data are available in dbGaP along with specifics of genotyping and QC (dbGaP: phs000718). After QC, genotypes were available for 2,186 unique study participants. We filtered SNPs based a deviation of Hardy-Weinberg equilibrium (HWE) at  $p < 10^{-6}$ . Overall, the median call rate was 99.86% and the error rate estimated from 53 pairs of study sample duplicates is  $1 \times 10^{-4}$ . All samples had a missing call rate  $< 2\%$ . The percent of SNPs with a minor allele frequency (MAF) of  $< 2\%$  was 30% for the autosomes and 32.1% for the X chromosome. This calculation was based on all study participants for SNPs not located on chromosome 21 and on only study parent samples for SNPs on chromosome 21. Trisomic genotypes for all 749 children in the study were called from raw genotyping data with previously-developed methods [17].

Possible chromosomal abnormalities beyond trisomy 21 were examined as possible artifacts of the use of DNA from LCLs. This was done using “Log R Ratio” (LRR) and “B Allele Frequency” (BAF) [39, 40] and applying the methods outlined in Laurie et al. [41]. Regions or chromosomes containing identified anomalies were excluded for genotype imputation purposes (see below). For chromosomes other than chromosome 21, 50 large anomalies were identified, of which 15 were filtered out of the dataset by setting genotypes in the identified region to missing. In addition, Mendelian inconsistencies were examined and one additional family was identified to have a genotype pattern consistent with uniparental chromosome 16 in the offspring. Genotypes at this chromosome were also set to missing.

Seven participants with neither parent genotyped were excluded from subsequent analyses. Thus in the remaining 742 families, genotypes were available for both the child and either the mother only ( $n = 95$ ), the father only ( $n = 17$ ), or both parents ( $n = 630$ ).

### **2.4.1.3 Adjustment for Population Structure**

Binary trait analyses using logistic regression are our primary statistical approach in this GWAS study. To adjust for population structure, we first used principal components analysis (PCA) as described by Patterson et al. [42], and implemented in R (SNPRelate package). SNPs used for PCA were selected by LD pruning from an initial pool that included all non-chromosome 21 autosomal SNPs with a missing call rate  $< 5\%$  and MAF  $> 5\%$ . In addition, the 2q21 (LCT), HLA, 8p23, and 17q21.31 regions were excluded from the initial pool. The first three eigenvectors were used in subsequent analyses.

### **2.4.1.4 Imputation**

The UWGCC used IMPUTE2 software [43] to perform genotype imputation. Details of their methods and QC are available at dbGaP:phs000718. The worldwide reference panel of 1,092 samples from the 1000 Genomes Project's Phase I integrated variant set [44] was used for imputation. We included only imputed variants with a quality metric of  $\geq 0.3$ , as previously recommended [45].

### **2.4.1.5 Phenotyping**

Our primary association studies were based on mothers who had a live birth with full or mosaic trisomy 21 as determined by karyotype and then determined to be due to a maternal nondisjunction error based on the characterization of the chromosome 21 genotype contributions from parent to the child with trisomy 21. Genotypes were obtained from the Illumina HumanOmniExpressExome-8v1-2 array and from previously genotyped variants along chromosome 21 using both STRs and SNPs [11, 13] The groups based on maternal nondisjunction

errors were compared with fathers of the children with trisomy 21 who represent a random group of individuals from the population.

Methods for defining the type of nondisjunction errors are described in detail in our previous work (e.g., [11, 13]). Briefly, parental origin of the meiotic error (maternal or paternal) was first determined by establishing the contribution of informative parental chromosome 21 genotypes to the child with trisomy 21. In families with both parents genotyped and where the parent of origin was unambiguously confirmed to be the mother (the vast majority of these families), we defined the meiotic stage of origin. We scored the genotype at each informative SNP and STR marker on chromosome 21q as either reduced to homozygosity (R) or not (N), according to whether the mother transmitted two identical or two different alleles, respectively, to her child at that locus. The meiotic stage of nondisjunction (MI or MII) was called according to the zygosity at the loci most proximal to the centromere (N or R, respectively). In a few cases (n=7), MII nondisjunction was called on the basis of a single, well-genotyped R SNP nearest the centromere (followed by a series of N SNPs), but because of the dense SNP genotyping on the chip, stage was more typically supported by many markers.

In families with only one parent genotyped, a slightly different approach was required, as missing parental data led to some markers that are partially informative, but not dispositive of zygosity. Briefly, we considered the ratio of information in the SNPs near the centromere, and called each case as MI or MII depending on the ratio. The threshold for this ratio was selected by performing an experiment with the complete trios; for each complete trio, we masked the genotype of one parent, calculated the ratio described above, and found the cutoff that optimized the predictive accuracy.

Lastly, when all informative markers in the parent of origin were reduced to homozygosity along 21q, the origin of error was inferred to be a post-zygotic, mitotic error and the case was excluded from this study, consistent with previous studies [9]. However, we recognize that a proportion of these cases may be MII nondisjunction errors with no recombination.

## **2.4.2 Analysis**

### **2.4.2.1 Sample Size**

As described above, samples from 2,186 participants were genotyped for this study, comprising 749 children and 1,437 parents. Participants with unresolved identity swaps, probands (children) with neither parent genotyped, and mothers in cases of potentially mitotically-arising trisomy were excluded from GWAS. After this filtering, 705 mothers and 645 fathers were retained for analysis, comprising 612 complete parent-child trios. Meiotic stage of origin for trisomy was determined to be MI in 535 cases, MII in 157 cases, and was not determined in 13 cases. Sample sizes for our analysis groups are reported in Table 2.1.

### **2.4.2.2 Association Studies**

We performed five primary GWAS analyses (summarized in Table 1). The comparison for all mothers vs. fathers can identify maternal genetic factors influencing nondisjunction either in MI or MII (or, more powerfully, in both). As noted in the Introduction, some genetic factors affecting MI nondisjunction may be shared with MII nondisjunction. Comparison of MI-only or MII-only mothers with fathers can identify maternal genetic factors influencing MI nondisjunction or MII nondisjunction, respectively. All three of these comparisons will also detect maternally-derived variants affecting survival of the infant to term. We chose to use fathers within our own



study as controls rather than turning to external controls because of the significant problem with confounding (chip and study effects) that is introduced when cases genotyped in one study are compared to controls genotyped in another.

However, one risk of using fathers as controls is that in theory the three analyses that compare mothers to fathers may also identify spurious associations due to comparing females to males. We tested this by running a female vs. male GWAS in a large additional dataset and comparing our results to those. The dataset we used was a subset of the COHRA study [46]; this study targeted dental phenotypes, but participants were selected in a community-based setting without regard to phenotype. We used 456 male and 494 female unrelated self-identified white adults in order to have a sample size comparable to the current study. By using sex as the outcome measure in a sample that was unselected with regard to phenotype, this analysis gave us a set of results to compare to our trisomy dataset in order to determine whether any of our trisomy results might instead be male vs. female artifacts. The female vs. male analysis in the COHRA dataset did not result in any unusually significant differences ( $\lambda = 0.94$ ). None of the GWAS loci or candidate genes described in the results section appeared among the largest differences between males and females in the COHRA dataset. The Manhattan plot and QQ plots are provided in the Supporting Information (S1 Figure), as well as results from the COHRA analysis in our candidate genes (S1 Table).

The fourth comparison involves MI vs. MII mothers. This comparison has the potential to identify unique factors for MI or MII nondisjunction without confounding by trisomy 21 survival; that is, both groups of mothers had a live birth child with trisomy 21.

For the fifth analysis, we conducted a transmission disequilibrium test (TDT) [47]. This test examines the association between the *child's* genotype and the dual phenotype of

nondisjunction and survival to term. Our prior hypothesis is that this test is best for identifying fetal “survival genes.” If there is association between *maternal* genotype and either nondisjunction or survival, this test can in theory identify it, but the association would be weakened. We did not perform this test for the candidate genes, since they were chosen as candidates for nondisjunction, not for survival. For the nondisjoined chromosome 21, we used a trisomic TDT, previously developed by our group [48].

For all analyses except the TDT, we used the logistic regression model  $\text{logit}(p) = \text{SNP} + \text{PC1} + \text{PC2} + \text{PC3}$ , where *SNP* is encoded additively and *PC1*, *PC2*, and *PC3* are the first three principal components of ancestry. The X chromosome was not examined because our primary comparative analyses involved mothers vs. fathers.

For all analyses, we filtered out SNPs with  $\text{MAF} < 1\%$  or with extreme departure from HWE. Imputed SNPs with info score  $< 0.5$  were also excluded, and imputed genotypes called with less than 90% confidence were coded as missing. Analyses were performed with PLINK and R.

### **2.4.2.3 Maternal Age Effect**

Because of the strong maternal age effect in maternal chromosome 21 nondisjunction, it is important to consider how maternal age fits into the above analyses. Previous results from our group and others suggest not only different etiologies for MI and MII nondisjunction, but likely different etiologies in different age groups. Statistically, this would suggest a model that includes not only maternal age effect but also an age X genotype interaction term. However, since our primary analyses compare mothers to fathers, it is not possible to fit such a model (since fathers have no “maternal age”). The logical analysis, then, is to stratify by maternal age group, similar to the approach we took for the MI and MII subgroups. We performed several such analyses, but the

sample sizes were prohibitively small for interpretation. We elaborate further on this issue in the Discussion.

#### **2.4.2.4 Candidate Gene Analyses**

For candidate gene analyses, we examined a window of 60kb on each side of the gene or SNP. We used a statistical significance cutoff based on the method of Li and Ji [49], which calculates the equivalent number of independent SNPs in the region and applies a Bonferroni correction based on that number. Thus the candidate gene analyses are fully corrected for multiple testing at the level of each individual gene.

#### **2.4.2.5 Follow-up Analyses to Examine Top-Ranked GWAS Signals**

For follow-up analyses of signals of  $p < 10^{-5}$  for the GWAS, we performed literature searches on genes within 500kb. For each of those regions, LocusZoom plots were created in all analyses to identify common associations across analyses.

## **2.5 RESULTS**

### **2.5.1 Candidate Gene Association Studies**

We focused on two sets of candidate genes/regions: genes that function in early stages of meiosis and that have been associated with accurate chromosome segregation (n=24) and regions associated with human recombination genome-wide counts (n=8) [15]. The Bonferroni-corrected statistical significance cutoffs along with all results are shown in Table 2.2 and LocusZoom plots

are provided in Figures 2.1-2.5 and in the Supporting Information (S2-S9 Figures). Each row in Table 2.2 represents one candidate locus. Each column represents an analysis. For each cell in the table, the most significant association at the locus (not always unique) is reported. P-values significant after correcting for multiple testing are marked with an asterisk and highlighted. Note that for each analysis in each gene, Table 2.2 lists the most statistically significant result, so that the SNP that appears in a given gene is not necessarily the same in each analysis. More detailed results are shown in the Supporting Information (S2 Table).

**Table 2.2 Candidate gene results**

Locus	All mothers vs. fathers	MI mothers vs. fathers	MII mothers vs. fathers	MI mothers vs. MII mothers	Significance threshold
SYCP1	P=0.00238	P=0.00255	P=0.002	P=2.69e-05*	9.43e-04
SYCP2	P=0.017	P=0.00592	P=0.000735*	P=3.09e-05*	8.62e-04
SYCP3	P=0.0078	P=0.00748	P=0.00637	P=0.00151	9.43e-04
SYCE1	P=0.0336	P=0.022	P=0.0227	P=0.00222	1.35e-03
SYCE2	P=0.00146	P=0.000425*	P=0.0154	P=0.00655	1.14e-03
SYCE3	P=0.00324	P=0.000764	P=0.0053518	P=0.00337	5.38e-04
TEX12	P=0.00558	P=0.0219	P=0.013871	P=0.0271	1.19e-03
BLM	P=0.0225	P=0.0378	P=0.00601	P=0.0257	7.04e-04
DMC1	P=0.0184	P=0.00993	P=0.018619	P=0.0134	1.19e-03
EXO1	P=0.00924	P=0.0118	P=0.00111	P=0.00303	6.10e-04
HORMAD1	P=0.0152	P=0.0174	P=0.011377	P=0.00361	1.35e-03
HORMAD2	P=0.0113	P=0.0156	P=0.00155	P=0.00838	8.93e-04
MEI1	P=0.00655	P=0.0485	P=0.00951	P=0.0204	9.26e-04
MEI4	P=0.0147	P=0.0164	P=0.0283	P=0.021	6.33e-04
MLH1	P=0.0151	P=0.00297	P=0.0162	P=0.0427	1.14e-03
MLH3	P=0.00472	P=0.0107	P=0.0241	P=0.0207	1.35e-03
MND1	P=0.0273	P=0.013	P=0.000336*	P=9.6e-05*	8.77e-04
MSH5	P=0.00677	P=0.00135	P=0.0417	P=0.0241	9.26e-04
REC114	P=0.0187	P=0.0272	P=0.0886	P=0.0118	8.77e-04
REC8	P=0.00822	P=0.00404	P=0.00195	P=0.00112	9.43e-04
SMC1B	P=0.0115	P=0.00407	P=0.047	P=0.0349	7.81e-04
SPO11	P=0.0232	P=0.0325	P=0.0115	P=0.0181	1.25e-03
STAG3	P=0.005	P=0.00472	P=0.0798	P=0.00958	1.19e-03
TRIP13	P=0.0026	P=0.0122	P=0.00358	P=0.0206	1.61e-03
rs1254319 (C14orf39 missense)	P=0.0278	P=0.0488	P=0.0372	P=0.00464	1.39e-03
rs75502650 (CCDC43 intron)	P=0.00203	P=0.00137*	P=0.000628*	P=0.031	7.58e-04
rs1132644 (CCNB1IP1 UTR)	P=0.0232	P=0.00657	P=0.0236	P=0.0195	1.32e-03
rs56162163 (chr17 inversion)	P=0.00666	P=0.00956	P=0.00737	P=0.0166	6.67e-04

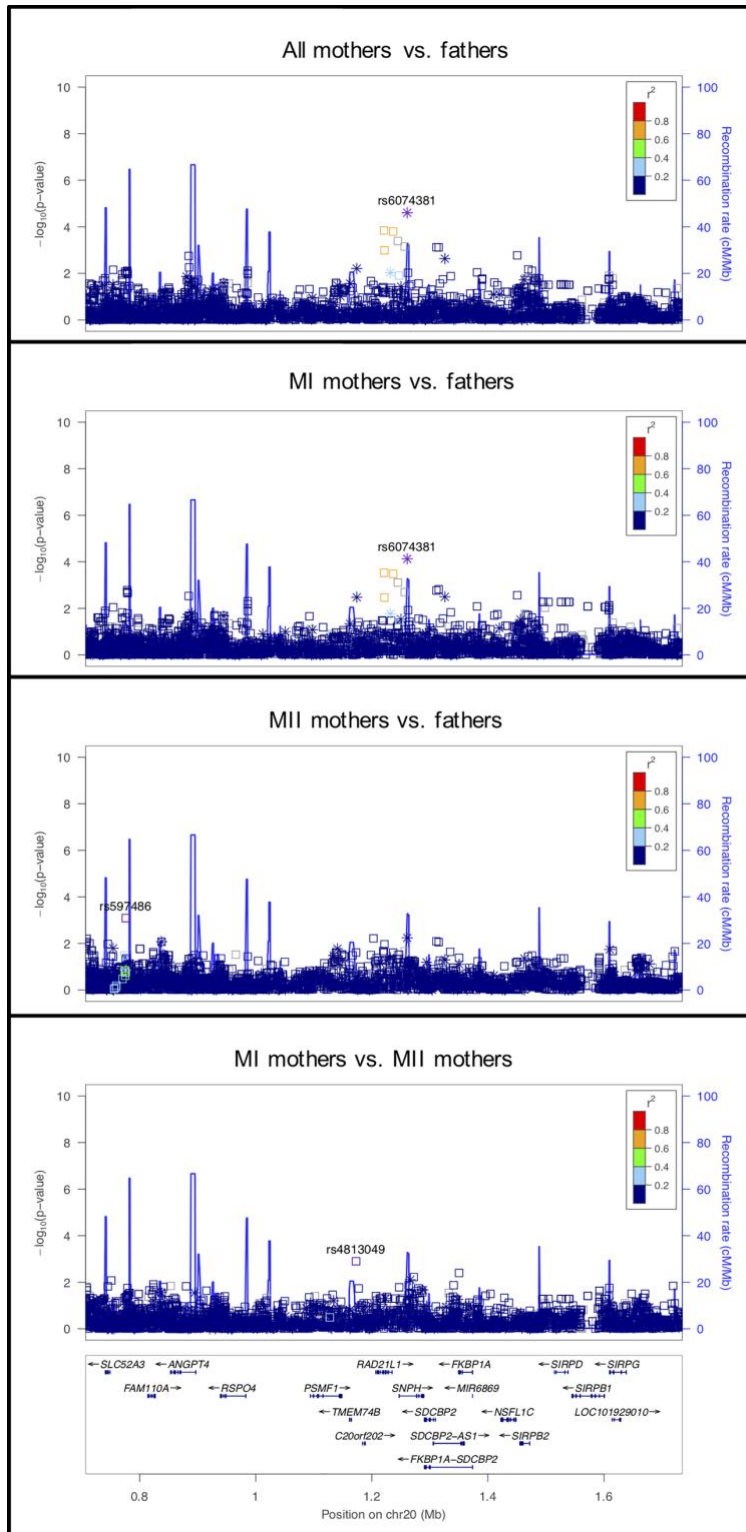
rs74434767 (CPLX1 intron)	P=0.00611	P=0.0178	P=0.0227	P=0.0329	1.61e-03
rs5745459 (MSH4 missense)	P=0.0389	P=0.0283	P=0.037	P=0.0198	1.16e-03
rs150798754 (PRDM9 intergenic)	P=0.00219	P=0.006	P=0.0123	P=0.022	1.25e-03
rs6889665 (PRDM9 upstream)	P=0.00219	P=0.006	P=0.0104	P=0.0678	8.33e-04
rs450739 (RAD21L missense)	P=2.49e-05*	P=7.47e-05*	P=0.00579	P=0.00126	1.09e-03
rs4045481 (RNF212 missense)	P=0.00475	P=0.0136	P=0.00514	P=0.0131	1.02e-03
rs658846 (RNF212 intron)	P=0.00292	P=0.0108	P=0.00514	P=0.0131	8.33e-04
rs12233733 (RNF212 nearby)	P=0.00292	P=0.0108	P=0.00396	P=0.031	1.52e-03
rs10135595 (SMEK1 UTR)	P=0.0125	P=0.007	P=0.00306	P=0.000183*	1.39e-03

Each row represents one candidate locus (either a gene with a 60kb border on each side or a 60kb window around a SNP). Each column represents an analysis. For each locus-analysis pair, the most significant association at the locus (not always unique) is reported. P-values significant after correcting for multiple testing (i.e., exceeding the Bonferroni-adjusted significance threshold noted in the last column) are marked with an asterisk and highlighted. (MI: meiosis I; MII: meiosis II; P: p-value; OR: odds ratio.) The first 24 loci represent genes selected for their function (above the double line). The latter 13 loci represent SNPs identified by Kong et al. in their GWAS of recombination 33, with annotation in parentheses (below the double line).

### 2.5.1.1 Candidate Genes Associated with Chromosome Segregation

These genes are shown in the top half (above the double line) of Table 2.2. Examination of genes involved in the meiotic cohesion complex showed a statistically significant association with *RAD21L*, a meiosis-specific member of the  $\alpha$ -kleisin protein family [50-53]. This was significant in both the mothers vs. fathers and the MI mothers vs. fathers comparisons, and has a similar effect (odds ratio) in the MII cases at the same SNP (Figure 2.1). Meiotic cohesins are essential for sister chromatid cohesion, but also have an effect on other prophase I processes, including formation of the axial/lateral elements, assembly of the SC, and crossing-over (e.g., [54, 55]). Gene disruption of *RAD21L* leads to sexually dimorphic phenotypes in mice. Male mice are infertile, whereas female mice show age-related infertility, reminiscent of primary ovarian

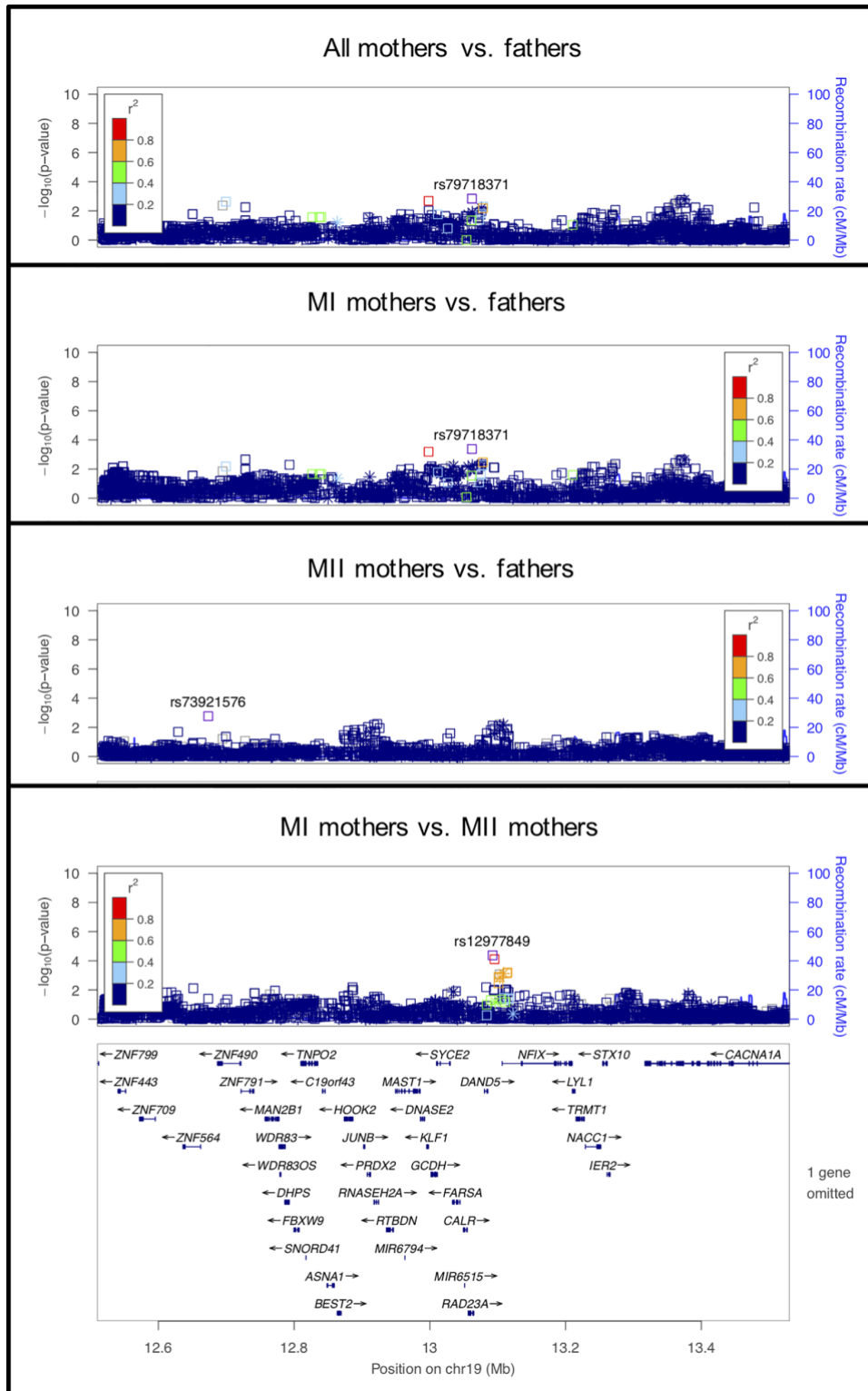
insufficiency. The reduced efficiency in synapsis in fetal oocytes may result in a lower ovarian reserve to be established [53] In human males, variants in RAD21L have been implicated in meiotic arrest and Sertoli cell-only syndrome [56].



**Figure 2.1 LocusZoom plot for *RAD21L***

Variants in seven genes coding for components of the SC were also investigated in this candidate gene group. Of the genes coding for components of the central element of the SC (i.e., *SYCE1*, *SYCE2*, *SYCE3*, *TEX12*), *SYCE2* showed a statistically significant association (in the MI mothers vs. fathers) (Figure 2.2), although the association with *SYCE3* was close to the cutoff for significance (also in MI mothers vs. fathers).





**Figure 2.2 LocusZoom plot for *SYCE2***

The other SC genes we examined code for the transverse filament (*SYCP1*) and components of the axial/lateral elements (*SYCP2* and *SYCP3*). *SYCP1* showed significant

association in the MI vs. MII analysis, but not in the other analyses (Figure 2.3). The signal in *SYCP1* was primarily located at an imputed SNP, at rs35401563, so this result requires confirmation by further genotyping. *SYCP2* showed highly significant associations in both the MII mothers vs. fathers and the MI vs. MII comparisons (Figure 2.4), suggesting the potential for an effect specific to MII. *SYCP3* was nearly significant in the MI vs. MII comparison.

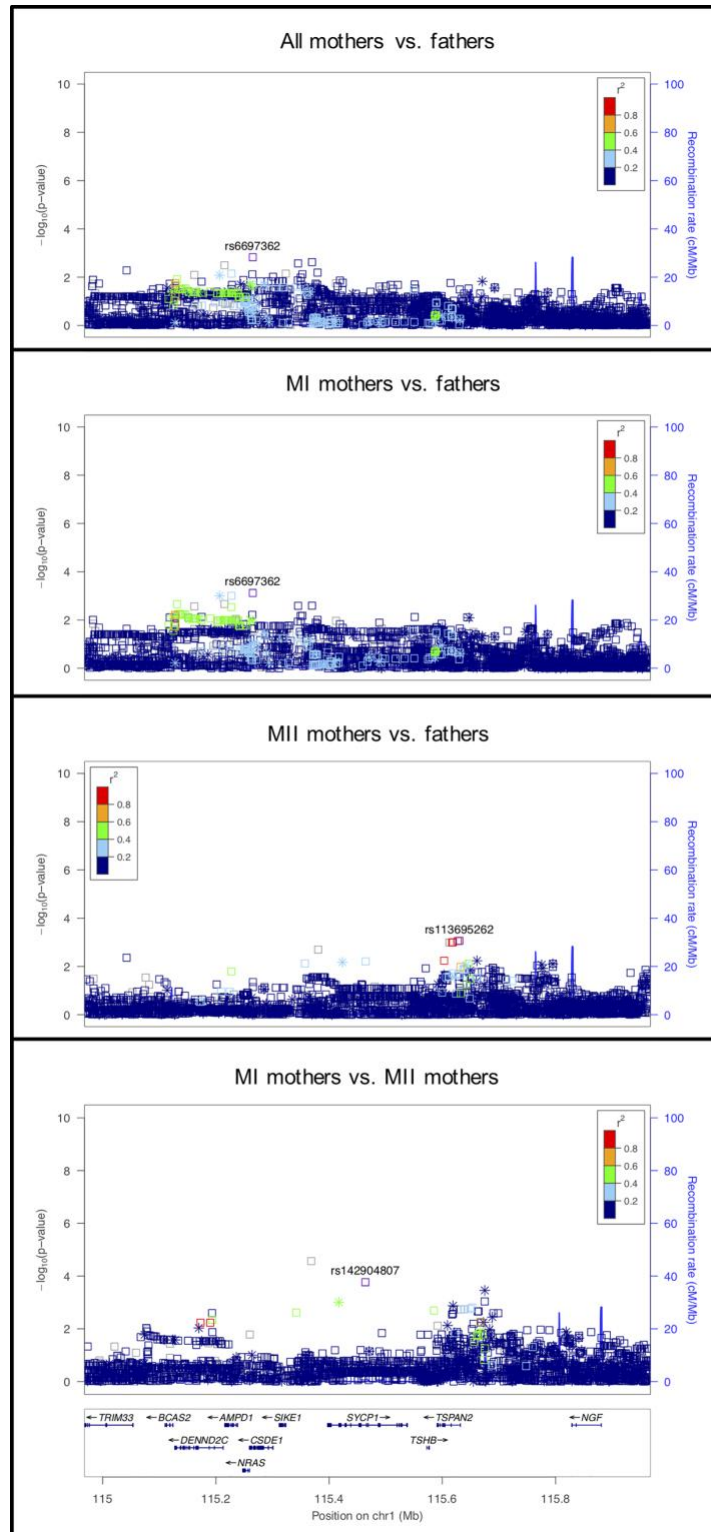
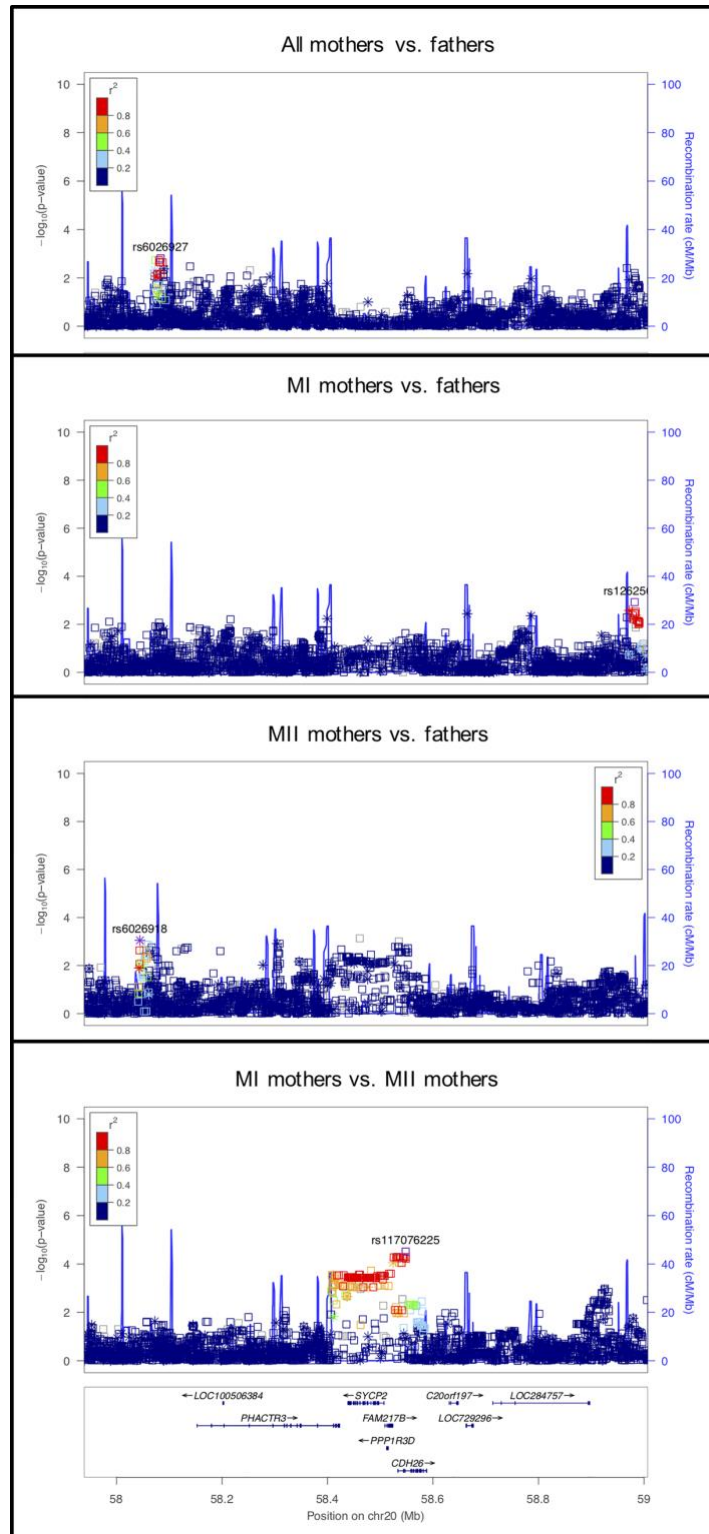


Figure 2.3 LocusZoom plot for SYCP1



**Figure 2.4 LocusZoom plot for SYCP2**

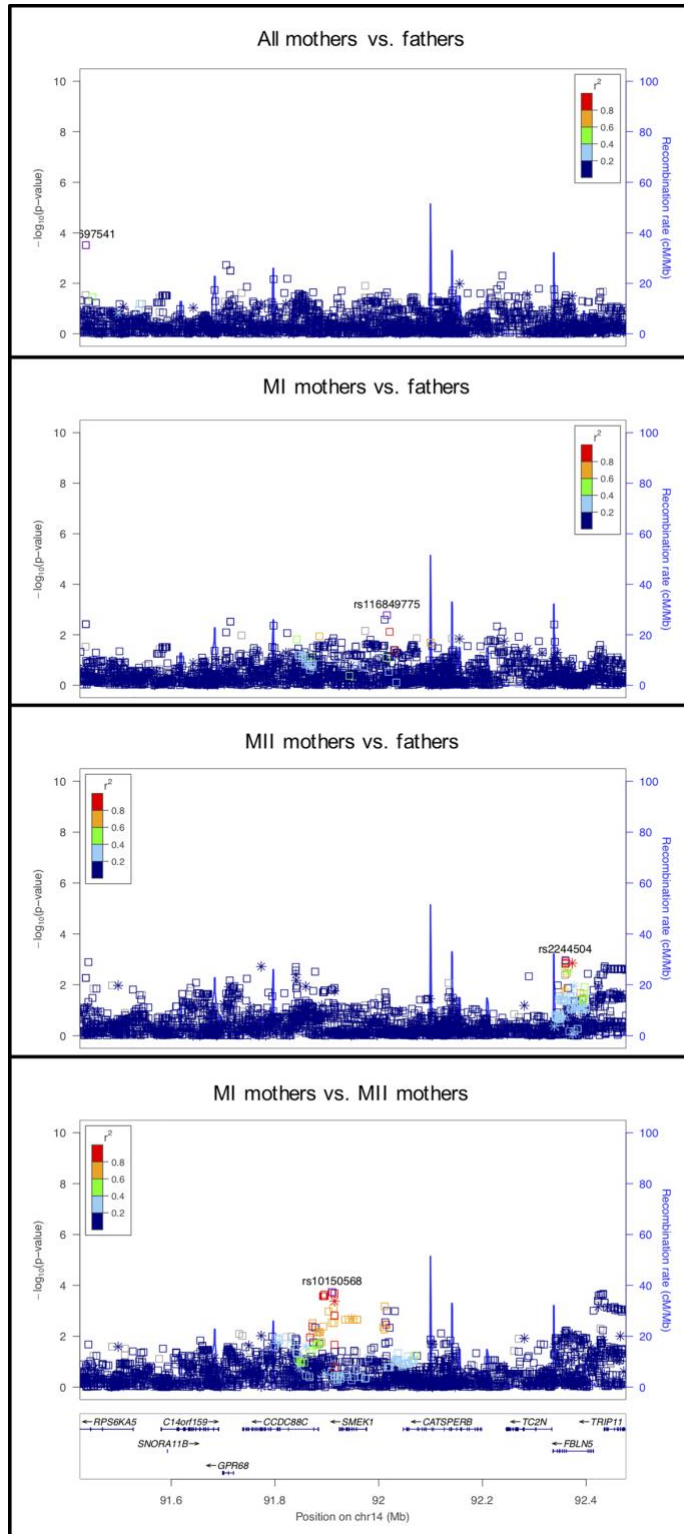
Among the other candidate genes in this group, the only statistically significant result was for *MND1*. The observed significant association was strongest in the MI vs. MII comparison, was

also strong in the MII vs. fathers comparison, and was much weaker in the MI vs. fathers comparison. This pattern suggests that this locus may be associated with MII nondisjunction. Based on genetic and cellular analysis of deletion mutants, MND1, acting with HOP2, plays a role in the initial processing of DSBs. Specifically, the HOP2-MND1 complex is involved in two separate stages of the DMC1-promoted recombination process: first, in the stabilization of DMC1 filaments on the resected end of the DSBs, and second, in the promotion of the subsequent strand invasion steps. In higher eukaryotes (mouse [57] and *Arabidopsis thaliana* [58, 59]), MND1 appears to be required for normal male and female fertility. Mutations result in normal recombination initiation, but meiotic DSBs are abnormally repaired and chromosome synapsis is aberrant [29]. The HOP2-MND1 complex has also been implicated in ovarian dysfunction and biochemically, is capable of driving RAD51-mediated alternative lengthening of telomeres in somatic cells [60]. If this association is confirmed, understanding why the effect of the variant is stronger in MII errors vs. MI errors may shed more light on its function.

#### **2.5.1.2 Genes Association with Human Genome-Wide Recombination Counts (Shown in the Bottom Half of Table 2.2, Below the Double Line)**

We also examined the eight regions identified in Kong et al. that were highly associated with genome-wide recombination counts in a large Icelandic study of 71,929 parent-offspring pairs [15]. Of the eight regions, three showed associations with maternal nondisjunction that were statistically significant according to the cutoffs shown in Table 2.2. The first region included *RAD21L* for which results are discussed above, as it was also in the group of candidate genes for meiotic processes. The second statistical signal was in the region of *SMEK1* (also known as protein phosphatase 4 regulatory subunit 3 (*PPP4R3A*)) and was strongest in the MI vs. MII mothers comparison (Figure 2.5). *SMEK1* is known as a regulator of cellular functions, including

apoptosis, cell growth, microtubule organization, cell cycle arrest, and TNF and PI3K/Akt signaling (e.g., [61, 62]). It is also known to play a role in endothelial cell function and subsequent angiogenesis [63]. However, its role in meiosis is unknown, although it is known to be expressed in the ovary.



**Figure 2.5 LocusZoom plot for *SMEK1***

The third signal is in the region of *CCDC43*, and was evident in both the MI mothers vs. fathers analysis and the MII mothers vs. fathers analysis (see S7 Figure). There is no known

function of CCDC43 in meiosis. In the study of Kong et al. [15], the SNP associated with recombination (rs75502650) was located in an intron of CCDC43. It was estimated to increase the global recombination rate by 76 cM and this effect was limited to females.

### 2.5.2 Strongest Results from the Genome-Wide Association Study

Because of the limited sample size in this study, the full GWAS produced only suggestive results, though a few of those top results have strong support in the literature for the relevance of the gene functions to meiosis or fetal survival. Manhattan plots and Q-Q plots for each GWAS analysis are included in the Supporting Information (S10 Figure). Tables 2.3-2.7 show the most statistically significant results from each of the comparisons in the genome-wide association study. For each result for a given comparison, the corresponding table also gives the smallest p-value within 20kb in each of the other comparisons. Detailed results are included in the Supporting Information (S3-S7 Tables).

**Table 2.3 Top hits from the all mothers vs. fathers genome-wide association study**

<b>Locus</b>	<b>All mothers vs. fathers</b>	<b>MI mothers vs. fathers</b>	<b>MII mothers vs. fathers</b>	<b>MI mothers vs. MII mothers</b>	<b>TDT</b>
rs10948101	P=7.65e-07	P=1.58e-06	P=0.000897	P=0.0634	P=0.00514
rs35141718	P=2.02e-06	P=1.92e-05	P=0.0109	P=0.127	P=0.0499
rs11535058	P=2.42e-06	P=2.1e-05	P=7.69e-05	P=0.0411	P=0.015
rs62086686	P=4.45e-06	P=0.000292	P=8.53e-05	P=0.0171	P=0.0133
rs75733466	P=5.09e-06	P=0.00023	P=1.24e-05	P=0.00189	P=0.0212
rs117746305	P=5.14e-06	P=0.000168	P=3.93e-06	P=0.056	P=0.0881
rs12947774	P=5.62e-06	P=0.000199	P=0.000949	P=0.00229	P=0.0339



rs1612273	P=5.93e-06	P=7.22e-05	P=1.62e-05	P=0.000934	P=0.00432
rs12652455	P=6.43e-06	P=2.43e-05	P=0.000437	P=0.0429	P=0.0518
rs148846406	P=6.79e-06	P=3.85e-05	P=0.00673	P=0.0562	P=0.00883
rs11026040	P=8.22e-06	P=0.000238	P=3.1e-05	P=0.00505	P=0.00794
rs7010571	P=8.27e-06	P=3.41e-05	P=0.0185	P=0.0796	P=0.0477
rs35816728	P=9.96e-06	P=6.09e-05	P=0.00101	P=0.00124	P=0.00815
Suggestive associations ( $p < 10^{-5}$ ) are recorded (highlighted cells). For each such locus, the most significant association within 20kb is recorded for each of the other four genome-wide analyses. Rows are ordered by significance. (MI: meiosis I; MII: meiosis II; P: p-value.)					

**Table 2.4 Top hits from the MI mothers vs. fathers genome-wide association study**

<b>Locus</b>	<b>All mothers vs. fathers</b>	<b>MI mothers vs. fathers</b>	<b>MII mothers vs. fathers</b>	<b>MI mothers vs. MII mothers</b>	<b>TDT</b>
rs10948100	P=7.65e-07	P=1.58e-06	P=0.000897	P=0.0634	P=0.00514
rs35288347	P=1.26e-05	P=2.72e-06	P=0.023	P=0.0298	P=0.0339
rs4649043	P=0.000148	P=3.1e-06	P=0.00388	P=0.000408	P=0.00556
rs437933	P=1.79e-05	P=3.29e-06	P=0.0683	P=0.05	P=0.0186
rs16847735	P=2.44e-05	P=3.77e-06	P=0.0173	P=0.0135	P=0.0617
rs2467011	P=5.02e-05	P=4.29e-06	P=0.0902	P=0.00786	P=0.158
rs9442389	P=1.88e-05	P=5.27e-06	P=0.0701	P=0.0164	P=0.0241
rs731245	P=0.000148	P=6.91e-06	P=0.00772	P=0.000809	P=0.00654
rs984968	P=1.48e-05	P=6.95e-06	P=0.0741	P=0.0425	P=0.0833
rs9984132	P=5.86e-05	P=9.87e-06	P=0.0539	P=0.00662	NA
Suggestive associations ( $p < 10^{-5}$ ) are recorded (highlighted cells). For each such locus, the most significant association within 20kb is recorded for each of the other four genome-wide analyses. Rows are ordered by significance. (MI: meiosis I; MII: meiosis II; P: p-value.)					

**Table 2.5 Top hits from the MII mothers vs. fathers genome-wide association study**

<b>Locus</b>	<b>All mothers vs. fathers</b>	<b>MI mothers vs. fathers</b>	<b>MII mothers vs. fathers</b>	<b>MI mothers vs. MII mothers</b>	<b>TDT</b>
rs1855111	P=6.48e-05	P=0.00213	P=2.2e-06	P=0.00698	P=0.00494
rs76740710	P=7.19e-06	P=0.000236	P=3.93e-06	P=0.0341	P=0.00648
rs12981234	P=0.0013	P=0.00911	P=4.28e-06	P=8.8e-05	P=0.0854
rs200216460	P=0.000209	P=0.0104	P=4.99e-06	P=0.00373	P=0.00284
rs11668205	P=0.00588	P=0.00218	P=5.49e-06	P=3.34e-05	P=0.0233
rs146838878	P=0.0197	P=0.0249	P=6.16e-06	P=0.000328	P=0.000463
rs115281615	P=0.047	P=0.0792	P=6.34e-06	P=7.33e-07	P=0.0712
rs62359711	P=0.0024	P=0.00108	P=6.91e-06	P=0.000191	P=0.0112
rs73178888	P=0.00754	P=0.0041	P=7.05e-06	P=1.72e-05	P=0.0122
rs9966603	P=0.00967	P=0.0639	P=7.05e-06	P=0.00208	P=0.00759
rs13020106	P=0.00149	P=0.0335	P=7.12e-06	P=0.00336	P=0.0162
rs2560850	P=0.0016	P=0.00365	P=7.69e-06	P=0.000332	P=0.00604
rs1191234	P=0.00176	P=0.0137	P=9.98e-06	P=0.00157	P=0.0411

Suggestive associations ( $p < 10^{-5}$ ) are recorded (highlighted cells). For each such locus, the most significant association within 20kb is recorded for each of the other four genome-wide analyses. Rows are ordered by significance. (MI: meiosis I; MII: meiosis II; P: p-value.)

**Table 2.6 Top hits from the MI mothers vs. MII mothers genome-wide association study**

<b>Locus</b>	<b>All mothers vs. fathers</b>	<b>MI mothers vs. fathers</b>	<b>MII mothers vs. fathers</b>	<b>MI mothers vs. MII mothers</b>	<b>TDT</b>
rs115281615	P=0.047	P=0.0792	P=6.34e-06	P=7.33e-07	P=0.0712
rs6440985	P=0.0958	P=0.0112	P=0.000198	P=1.31e-06	P=0.00796
rs2806747	P=0.0934	P=0.0656	P=0.00037	P=1.35e-06	P=0.00284
rs9319652	P=0.0225	P=0.0267	P=0.00243	P=4.09e-06	P=0.039

rs11977478	P=0.00474	P=0.00902	P=0.00019	P=4.18e-06	P=0.00232
rs11034351	P=0.0124	P=0.00284	P=7e-04	P=4.3e-06	P=0.0295
rs7685548	P=0.00618	P=0.000955	P=0.00884	P=4.82e-06	P=0.00226
rs71967233	P=0.044	P=0.0785	P=0.000135	P=5.24e-06	P=0.0163
rs34282937	P=0.0334	P=0.0368	P=5.95e-05	P=5.48e-06	P=0.0298
rs77525287	P=0.0707	P=0.000444	P=0.0124	P=5.9e-06	P=0.099
rs4818884	P=0.0454	P=0.0312	P=0.00354	P=9.57e-06	NA
rs61999085	P=0.00851	P=0.000125	P=0.0449	P=9.84e-06	P=0.00243

Suggestive associations ( $p < 10^{-5}$ ) are recorded (highlighted cells). For each such locus, the most significant association within 20kb is recorded for each of the other four genome-wide analyses. Rows are ordered by significance. (MI: meiosis I; MII: meiosis II; P: p-value.)

**Table 2.7 Top hits from the TDT (transmission disequilibrium test)**

<b>Locus</b>	<b>All mothers vs. fathers</b>	<b>MI mothers vs. fathers</b>	<b>MII mothers vs. fathers</b>	<b>MI mothers vs. MII mothers</b>	<b>TDT</b>
rs3802065	P=0.0144	P=0.0066	P=0.0318	P=0.00859	P=4.84e-07
rs2867076	P=0.0164	P=0.0227	P=0.008	P=0.00904	P=1.05e-06
rs7451700	P=0.0186	P=0.0169	P=0.054	P=0.0195	P=1.71e-06
rs7389783	P=0.0421	P=0.0397	P=0.0048	P=0.00741	P=1.93e-06
rs17769147	P=0.182	P=0.152	P=0.0738	P=0.0624	P=2.06e-06
rs201634098	P=0.00592	P=0.00385	P=0.0722	P=0.015	P=3.52e-06
chr23:154539980	NA	NA	NA	NA	P=3.63e-06
rs74615884	P=0.0801	P=0.037	P=0.0167	P=0.00236	P=3.73e-06
rs158866	P=0.036	P=0.00582	P=0.0103	P=0.0102	P=3.77e-06
rs1187600	P=0.0553	P=0.0636	P=0.204	P=0.156	P=6.25e-06
rs55743346	P=0.0514	P=0.0371	P=0.0834	P=0.0234	P=7.34e-06
rs183199067	P=0.0131	P=0.00285	P=0.0396	P=0.0254	P=9.55e-06
rs140022090	P=0.0163	P=0.00733	P=0.0538	P=0.00493	P=9.55e-06

rs34518363	P=0.00538	P=0.00186	P=0.049	P=0.0394	P=9.55e-06
rs6681167	P=0.0358	P=0.0505	P=0.0648	P=0.0377	P=9.58e-06
Suggestive associations ( $p < 10^{-5}$ ) are recorded (highlighted cells). For each such locus, the most significant association within 20kb is recorded for each of the other four genome-wide analyses. Rows are ordered by significance. (MI: meiosis I; MII: meiosis II; P: p-value.)					

### 2.5.2.1 rs10948101 on chromosome 6 near *VEGFA*

The observed signal for this locus was strongest for the analysis of all mothers vs. fathers and was located within *LOC100132354*, a long non-coding RNA (lncRNA) (Figure 2.6). Upstream of this intergenic lncRNA is *VEGFA*, the gene encoding vascular endothelial growth factor A (VEGFA). In a recent meta-analysis, *LOC100132354* was confirmed to be highly associated with VEGF circulating levels in serum [64]. VEGFA plays multiple roles in ovarian development and function (reviewed in McFee and Cupp [65]). Vascularization plays a role in the formation of early ovarian structures, primordial follicle assembly, and follicle activation. Further, ovarian function is highly dependent on the development and continual remodelling of a complex vascular system. This allows the follicle and the corpus luteum to receive the needed oxygen, nutrients, and systemic hormones and the release of ovarian hormones (reviewed in Robinson et al. [66]). If angiogenesis is disrupted, follicular growth is reduced, ovulation is perturbed, and development and function of the corpus luteum is significantly altered. The action of VEGFA is necessary at all these stages of development. We did not see any significant effect in the TDT analysis, so there is no suggestion that this locus is associated with survival to term.

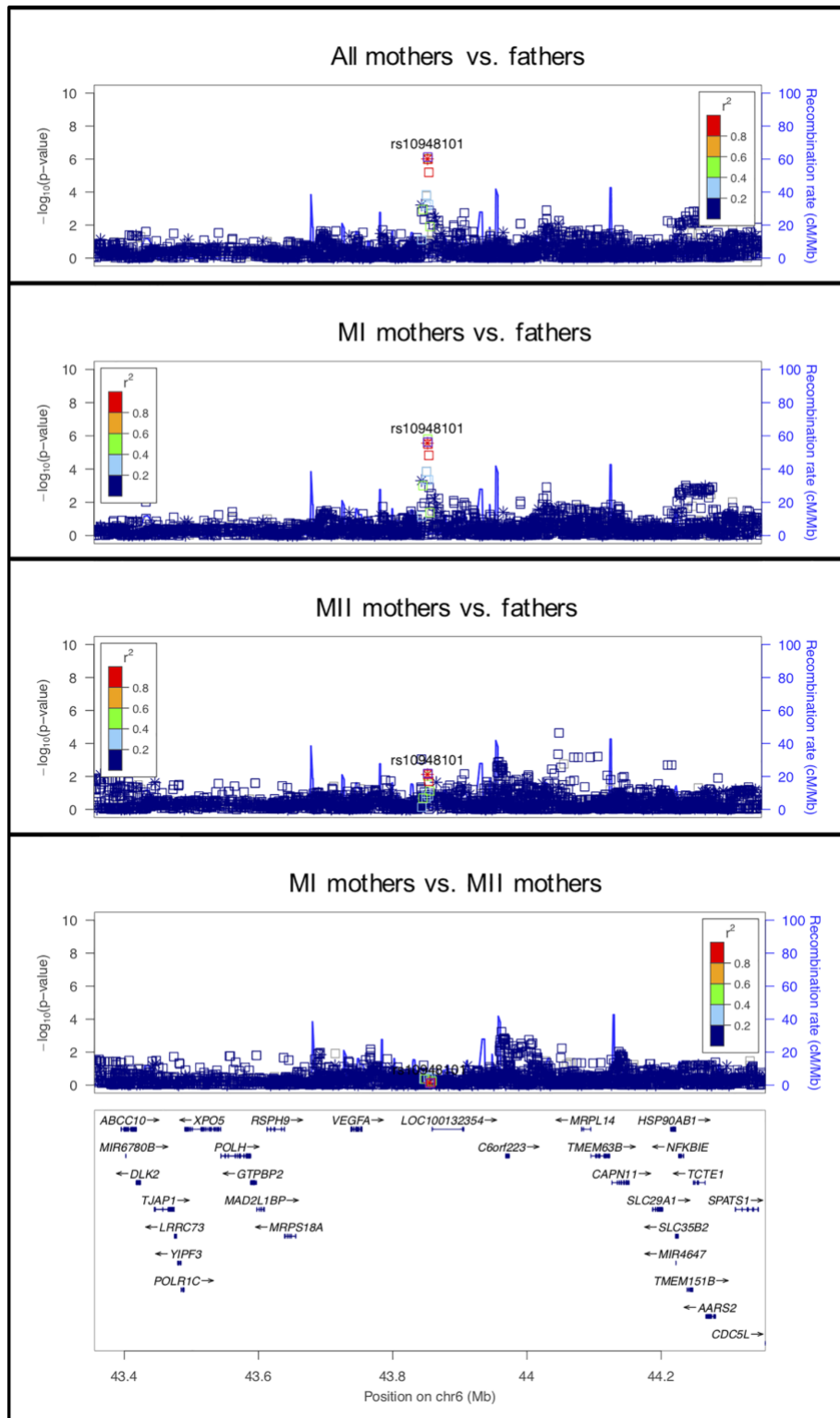
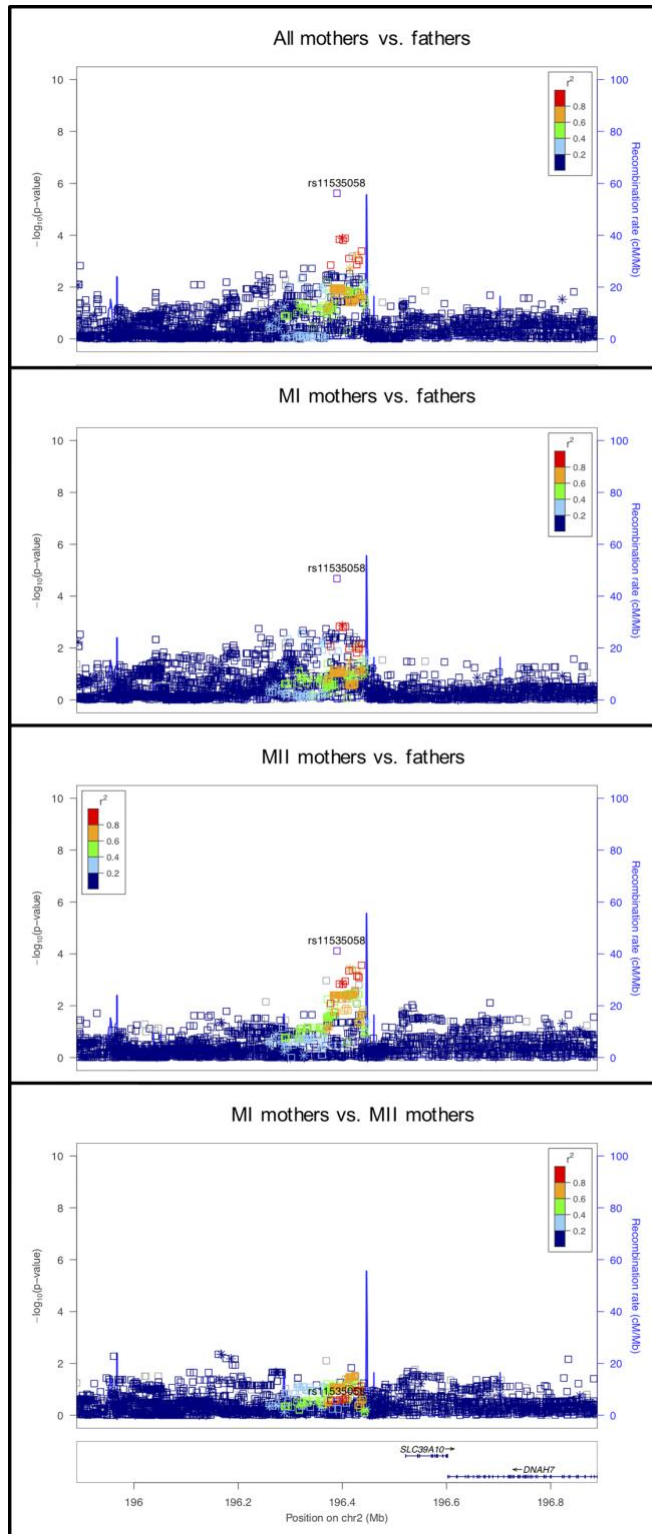


Figure 2.6 LocusZoom plot for *VEGFA* locus

### 2.5.2.2 rs11535058 on chromosome 2 near *SLC39A10*

The observed signal in this region, strongest in the mothers vs. fathers comparison, was located 132kb downstream of *SLC39A10* (Figure 2.7), with a similar, but non-significant, effect size for both MI mothers vs. fathers and MII mothers vs. fathers. *SLC39A10* is involved in the zinc transport network. Regulation of intracellular zinc is essential for oocyte maturation and activation. In mouse, progression of the oocyte from a cell arrested in prophase of MI into a mature egg arrested at metaphase of MII is accompanied by an increase in total zinc content. This increase is required for proper meiotic progression [67]. Also, exit from MII during oocyte activation requires decreasing cellular zinc through the rapid export of zinc from the oocyte. These ‘zinc sparks’ are required for oocyte activation and resumption of the cell cycle [68].

Specific to *SLC39A10*, Lisle et al. [69] found a complex zinc transport network present in the cumulus-oocyte complex in mouse oocytes. They found that mRNA transcripts for specific zinc transporter proteins (SLC family), including *Slc39a10* were higher in oocytes, while another unique set of zinc transporter protein transcripts were higher in cumulus cells. Thus, zinc homeostasis, regulated in the cumulus-oocyte complex, may affect both MI and MII processes.

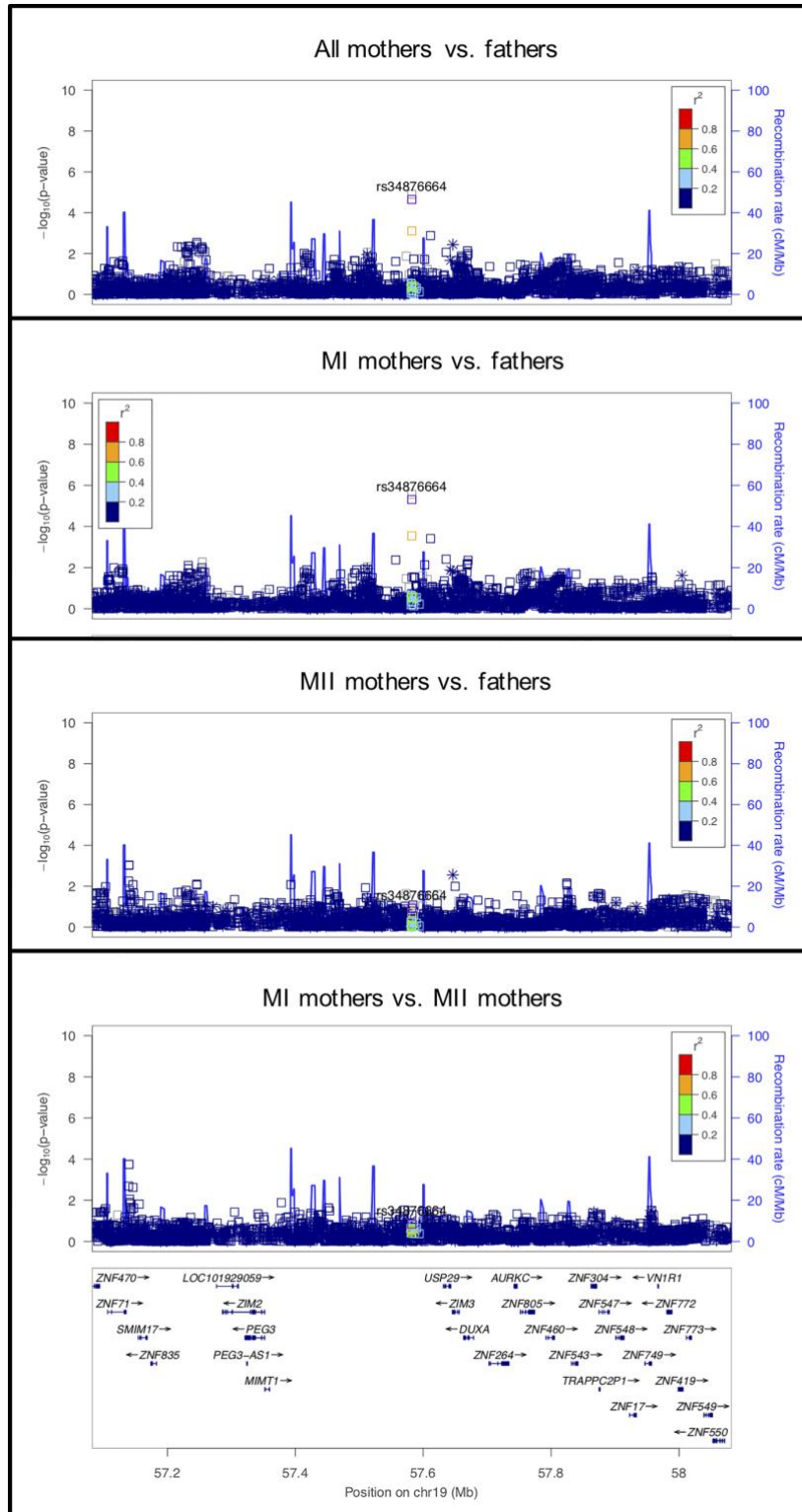


**Figure 2.7 LocusZoom plot for *SLC39A10* locus**

### **2.5.2.3 rs35288347 on chromosome 19 near *AURKC***

This signal occurs primarily in the MI mothers vs. fathers analysis (Figure 2.8), with a similar but less significant effect in the mothers vs. fathers analysis. Consistent with the possibility of an effect specific to MI, mutations in the *AURKC* gene in this region (160kb away) cause tetraploidy in human sperm and MI arrest in mouse oocytes [70, 71].





**Figure 2.8 LocusZoom plot for *AURKC* locus**

#### 2.5.2.4 rs9984132 on chromosome 21 located in a gene rich region

The signal at rs9984132 on chromosome 21 is located in a gene-rich region (Figure 2.9). The strongest signal at this locus was identified in the comparison of MI mothers with fathers. Two genes stand out as possible candidates for involvement in chromosome segregation. *COL6A2*, located 34kb upstream of the signal, codes for one of the components of collagen that is part of the extracellular matrix (ECM) formed by cumulus cells. This ovarian follicular ECM is related to proliferation, steroidogenesis, and luteinization [72]. As the formation of this ECM is involved in fertilization and embryo quality, the observation that this signal appears to be only related to MI nondisjunction reduces the support of *COL6A2* as a candidate.

*PCNT* is a gene located about 260kb upstream from the signal. Pericentrin, coded by *PCNT*, is a highly conserved component of the acentriolar microtubule-organizing centers (aMTOCs) in mouse oocytes. aMTOCs play a vital role in meiotic spindle assembly and stability. Depletion of pericentrin in mouse oocytes leads to increased rates of aneuploidy [73]. Human oocytes differ from mouse oocytes in that they lack *PCNT* and aMTOCs in MI, where spindle assembly is mediated from chromosomes by the small guanosine triphosphates [74]. Thus, more work is needed to confirm this signal and its underlying genetic association.

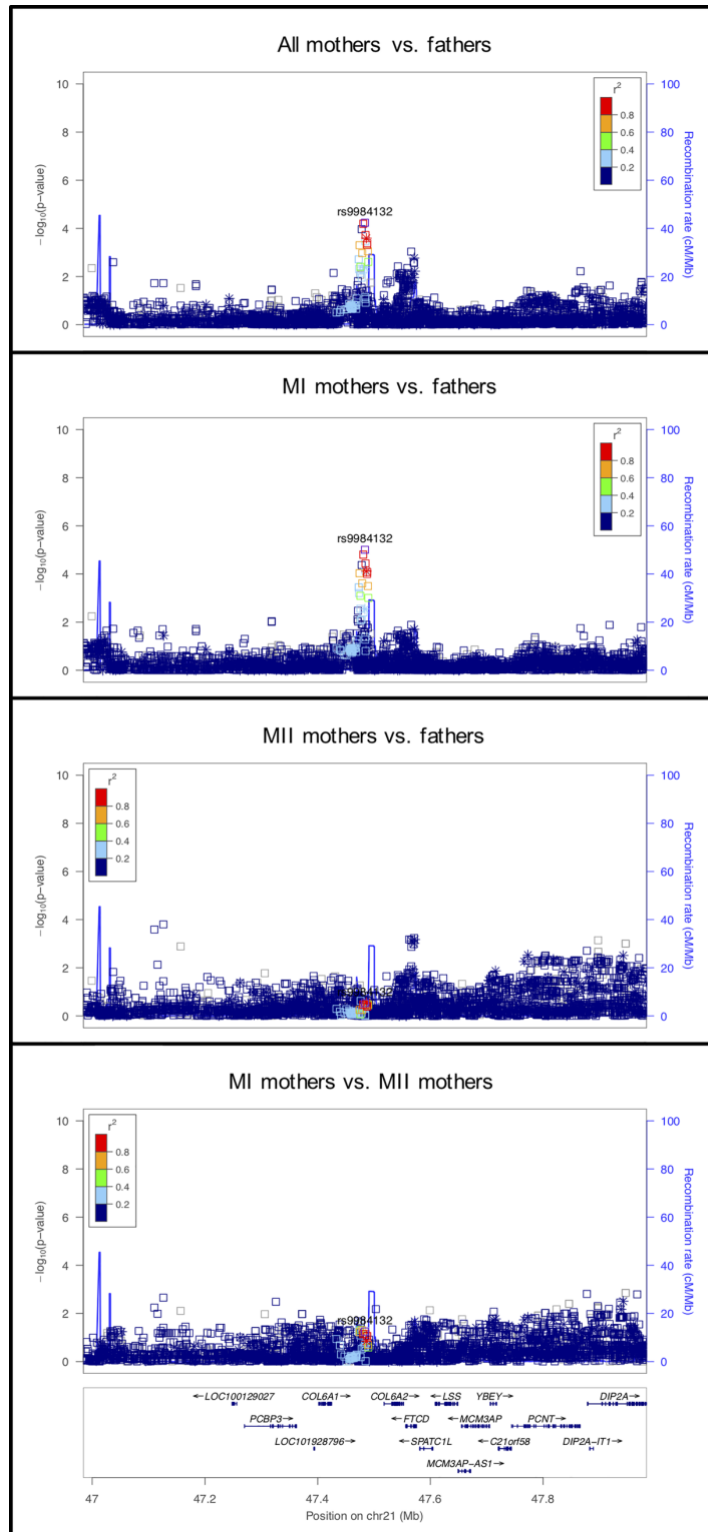


Figure 2.9 LocusZoom plot for rs9984132 locus

#### **2.5.2.5 rs73178888 on chromosome 8 near a region associated with meiotic recombination**

This signal, located in an intron of *ERICHI*, is primarily observed in the MII mothers vs. fathers analysis and in the MI vs. MII analysis, suggesting that it might be an MII risk locus (Figure 2.10). The location is noteworthy because Begum et al. [75] identified a variant 59kb away in this region as potentially associated with meiotic recombination (specifically recombination outside of hotspots) in a euploid population. There is no evidence to suggest that *ERICHI* or *DLGAP2* (also known as *ERICHI-ASI*), 3kb from the signal, is involved in recombination or meiosis. The next closest gene is *TDRP*. The deficiency of *TDRP* in mice is suggested to be involved in sperm motility and may play a role in spermatogenesis [76], but there is no evidence for involvement in oogenesis.

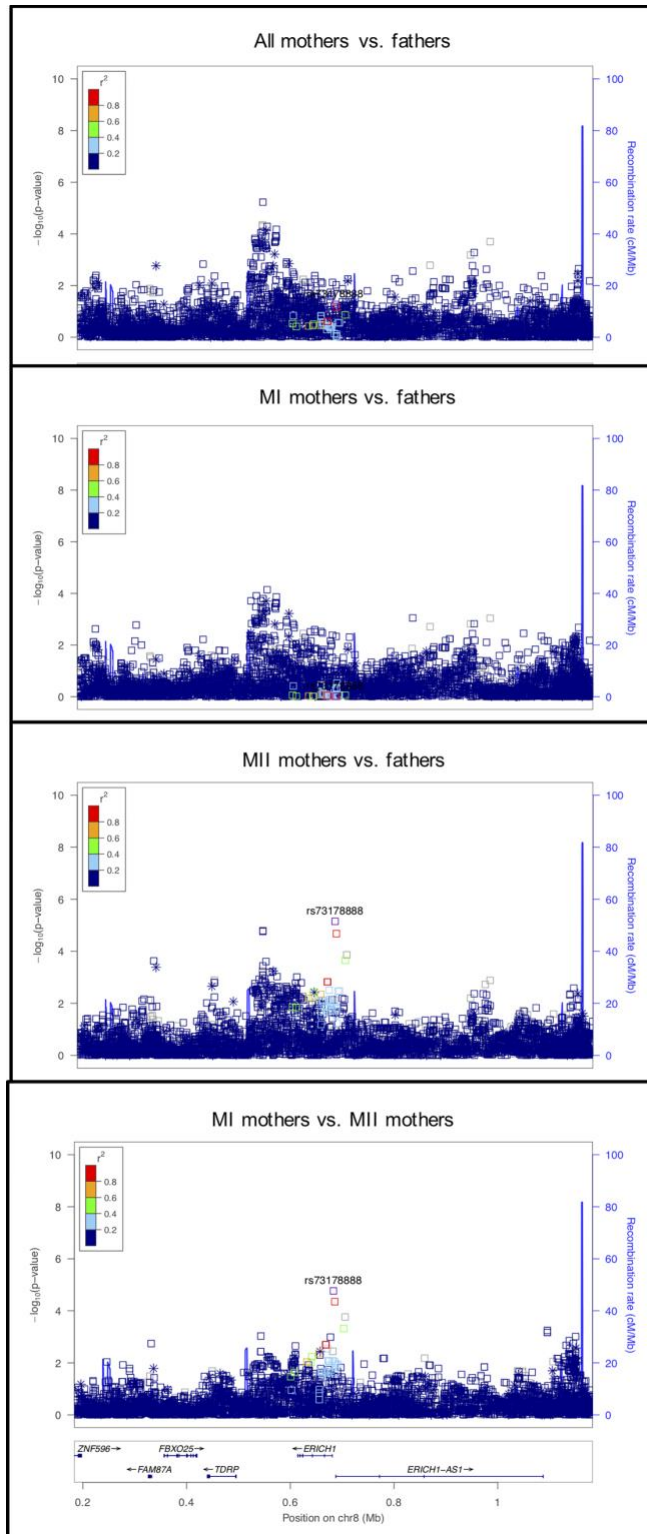
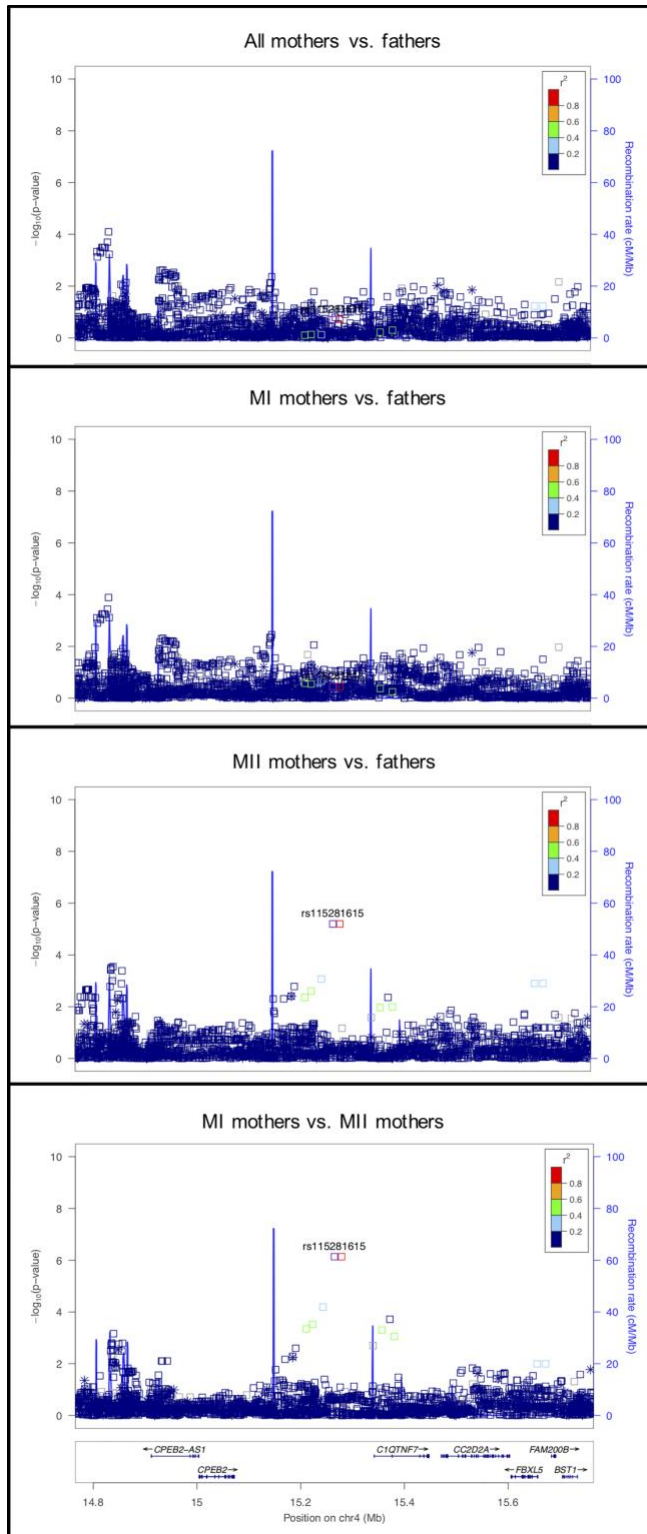


Figure 2.10 LocusZoom plot for *DLGAP2* locus

#### **2.5.2.6 rs115281615 on chromosome 4 near *CPEB2***

The signal at this locus is primarily observed in the MII mothers vs. fathers comparison (Figure 2.11). The genes in the region, *CIQTNF7*, and *CC2D2A*, do not have evidence for a role in meiosis. *CPEB2*, located 193kb from the signal, encodes an RNA-binding protein, cytoplasmic polyadenylation element binding protein and is thought to be involved in regulated translation, a system that allows the rapid production of selective proteins in response to a physiological signal. *CPEB2* interacts with the elongation factor, eEF2, to slow down peptide elongation of *CPEB2*-bound RNA [77]. In mice, this protein is highly similar to the family of *CPEBs* that regulate cytoplasmic polyadenylation of mRNA as a trans-factor in oogenesis and spermatogenesis. *CPEB2* is expressed post-meiotically in mouse spermatogenesis, which suggests a possible role in translational regulation of stored mRNAs in transcriptionally inactive haploid spermatids [78].

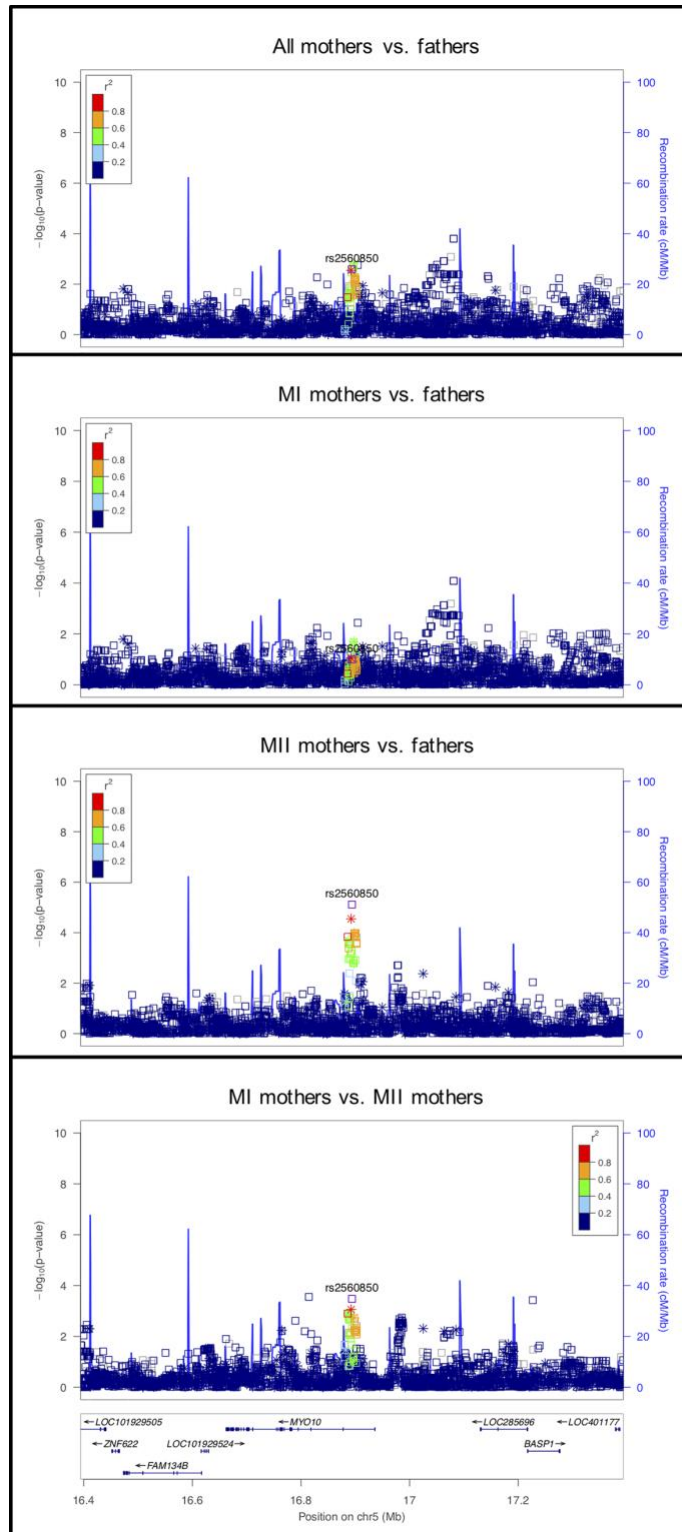


**Figure 2.11** LocusZoom plot for *CPEB2* locus

### **2.5.2.7 rs2560850 on chromosome 5 in an intron of *MYO10***

The comparison between MII mothers vs. fathers shows the strongest statistical significance for this locus which includes *MYO10* (Figure 2.12). Myosin-10 proteins are phosphoinositide-binding, actin-based motors that play an important role during meiosis in the integration of the F-actin and microtubule cytoskeletons. Proper spindle positioning and orientation are essential for asymmetric cell division and these functions are particularly important in meiosis. In *Xenopus* oocytes, Weber et al. [79] showed that myosin-10 is associated with microtubules and is concentrated where the meiotic spindle contacts the F-actin-rich cortex. This observation and others suggest that myosin-10, the microtubule-binding myosin, is required for anchoring the spindle and an actin-binding kinesin is required for meiotic cytokinesis [79, 80]. Recently F-actin was shown to permeate the microtubule spindles in oocytes of many mammals, including human, where it prevents lagging chromosomes and thus segregation errors, including during anaphase I [81].





**Figure 2.12** LocusZoom plot for *MYO10* locus

### 2.5.3 Strongest Results from the TDT Analyses

Together with the other case/control association tests to identify genes associated with nondisjunction, we also had the opportunity to use the TDT. We assumed that the TDT might help tease apart genes associated with nondisjunction from those associated with fetal “survival genes.” If there is association between *maternal* genotype and either nondisjunction or survival, this test can theoretically identify it. We only performed this test in our GWAS approach, not our candidate genes, since the candidates were chosen for possible involvement in nondisjunction per se. Here we highlight two genes where the statistical signal was relatively strong, although not genome-wide significant.

The first is located at rs17769147, 24kb upstream of the gene *RNF182*, which encodes a RING-finger-containing transmembrane protein that includes an E3 ubiquitin ligase activity. There was no statistical signal for any of the case/control comparisons, only for the TDT. Studies have shown that this gene is expressed preferentially in the brain and is up-regulated in Alzheimer’s disease brain and in neuronal cell cultures that are subject to stress-induced cell death [82]. In another line of study, *RNF182* was found to be one of the gene targets of MeCP2, the gene involved in Rett syndrome. The group of identified gene targets are involved in the regulation of the cell growth and survival of neuronal cells [83].

Another strong TDT signal was found at rs158866, with no statistical signal for any of the case/control comparisons. This signal is located within *NEDD4L* (also referred to as *NEDD4-2*), a gene encoding a ubiquitin ligase. This protein binds and regulates membrane-associated proteins (although not exclusively), particularly ion channels and transporters (reviewed in Goel et al. [84]). *NEDD4L* interacts with several other proteins and may regulate other important substrates as well

(e.g., [85, 86]). Based on current evidence, this ligase is essential for the maintenance of cellular homeostasis.

## 2.6 DISCUSSION

We present, for the first time, a candidate gene study and GWAS of chromosome 21 maternal nondisjunction. The goal of this project was to gain insight into factors that may predispose a woman to this common chromosomal error.

### 2.6.1 Genes Associated with Cohesin Complex

The meiosis-specific cohesin subunits are encoded by *SMC1 $\beta$* , *REC8*, *RAD21L*, and *STAG3*. Of these genes, we found that variation in *RAD21L* was associated with nondisjunction, with the strongest signal with MI nondisjunction. As part of the cohesion complex [50-53], *RAD21L* plays a role in the structural maintenance of chromosomes (SMC) complex. The SMC complex includes cohesin, condensin and SMC5/6, and is an important regulator of chromosome dynamics and structure during both mitosis and meiosis. In female mice, mutations in meiosis-specific cohesins and in the SMC complex increase the frequency of oocyte aneuploidy and primary ovarian insufficiency [53, 87-89]. The study of Kong et al. [15] found a highly significant association of *RAD21L* with male recombination and a much weaker signal in females; Begum et al. [75] did not replicate this finding, although the genomic region had poor coverage in two of the three datasets of their study. Together, this suggests a sex-specific role of *RAD21L* and one in

females that maybe more directly related to segregation of bivalents than recombination counts per se.

There are several possible explanations for why genetic variation in *RAD21L* was associated with maternal nondisjunction whereas variation in *REC8* was not, although both are  $\alpha$ -kleisin subunits that are part of the cohesin complex. One possible explanation is that there is reduced power to detect a signal based on allele frequencies of variants in *REC8* compared with *RAD21L*. Alternatively one gene may play a more essential role in meiosis, where variation is not tolerated or not compatible with the oocyte surviving to fertilization. It is known that these genes play unique roles during meiosis and thus have differential effects on the meiotic process [90-92].

### **2.6.2 Genes Associated with the Synaptonemal Complex (SC)**

The general structure of the SC is highly conserved across yeast and mammals, although the genes and proteins involved are not always conserved (reviewed in Cahoon and Hawley [22]). The tripartite protein structure extends along the entire length of the synapsed homologues and assembles alongside cohesin and cohesin-like proteins that hold the sister chromatids of the homologues together (e.g., [93, 94]). Mutations in genes coding for SC components have been identified previously among women with infertility or recurrent miscarriages (reviewed in Geisinger and Benavente [95]). At this time, only mutations in *SYCP3* and *SYCE1* have been identified, but most studies had <100 women available for study. In our study, we found evidence for an association of variants in genes of all three SC components with maternal nondisjunction. We had supporting evidence for *SYCP3* (although not statistically significant), with the strongest statistical signal being found in the comparison of MI vs. MII. In addition, we found a statistically significant association of *SYCE2* in MI vs. fathers and with *SYCP2* in the comparison of MI vs.

MII. Thus, with an increased sample size and a more homogeneous reproductive outcome, we were able to confirm the importance of the SC structure for proper segregation of chromosome during human oogenesis.

### **2.6.3 Association with Recombination-Related Variants**

It is now well established that there is both significant sex-specific and individual variation in genome-wide recombination counts and location of events, in spite of the need for the recombination process to be tightly controlled [96-98]. When there are alterations in the number of recombinants (reduced or no recombination) or their location (pericentromeric or telomeric), there is a high risk for human chromosome nondisjunction [99-101]. Variation in genes that play a role in recombination has been identified and we examined those that were identified in a large Icelandic study using linkage analyses of live births [15]. Begum et al. [75] also attempted to replicate the association of these variants in a population of primarily Northern European ancestry. Their GWAS meta-analyses were extended to the study of recombination phenotypes, including the average recombination count along with those related to placement relative to historical recombination hotspots. Here, we asked whether these variants would also explain susceptibility to maternal nondisjunction. Our results are interesting both with respect the identification of associated regions and to the lack of evidence in the others.

Both Kong et al. [15] and Begum et al. [75] found a strong association of *SMEK1* (also known as *PPP4R3A*) with recombination in females only. Our results identified the strongest association in the comparison of MI vs. MII nondisjunction, which suggests a stage-specific role of this protein, as well as a sex-specific role identified in the recombination studies. At this time, there is no known role of *SMEK1* in meiosis. It is a member of the PP2A subfamily. PP2A is

involved in de-protection of centromere cohesin in MII of mammals, a process that is essential for proper sister chromatid segregation (reviewed in Wassmann [102]). Although it is intriguing to think that variation in *SMEK1* may alter this MII-associated process, this is no direct link at this time.

Another locus that deserves follow-up is on chromosome 8 near rs73178888, one of the top hits in our genome-wide analyses. Begum et al. [75] identified this same region as potentially associated with meiotic recombination in a euploid population. Although the genes in the region do not appear to be linked to recombination, further investigation is warranted.

With respect to the two most well-established genes associated with recombination, *RNF212* and *PRDM9*, our data showed no association with nondisjunction. Variation in *RNF212* is sex-specific, some variants being associated with increased recombination in males and others in females [15, 75, 98, 103-105]. *RNF212* is known to form many discrete foci along chromosomes early in meiotic prophase I; these foci are then reduced to a few sites where crossovers are formed [106]. For small chromosomes such as chromosome 21, perhaps female-specific variation in this process is less evident compared with genome-wide alterations.

Variation in *PRDM9* is known to be associated with recombination hotspots in both males and females. Kong et al. [15] showed that variants were also associated with total recombination counts in males, but not females. Begum et al. [75] provided further evidence that in females, variants were associated with both recombinant counts within and outside of historical hotspots, in opposite directions. They suggested that females might have a compensatory mechanism, such that increased recombination in hotspots is balanced by decreased recombination elsewhere; thereby not altering the overall recombination count. In males, variants were only associated with recombinants within historical hotspots. In our previous study, we found that historical hotspot

usage along maternally-derived nondisjoined chromosomes 21 was similar to that in controls, particularly among MI errors, indicating that the observed altered telomeric placement probably does not involve differential hotspot usage [12]. Subsequently, Oliver et al. [107] studied sequence variation in the zinc finger-binding domain (ZFBD) of *PRDM9* in a subset of the study sample presented here. They found that the frequency of the *PRDM9* ZFBD minor alleles was significantly increased among women who had a chromosome 21 nondisjunction event and no observed recombination on 21q. Further, when these *PRDM9* minor alleles were compared with the major A-allele, fewer predicted binding sites on 21q were found. Together, these observations suggest that allelic variation in *PRDM9* may play a role in nondisjunction, but that the effect size may be small and it may be limited to nondisjunction of achiasmate chromosomes.

#### **2.6.4 Gene Discovery**

When we conducted a GWAS, no variants were genome-wide significant; thus, the marginally significant signals need replication prior to additional speculation. We highlighted a few findings in Results for signals in genes that are known to be involved in oogenesis. If these are true signals, our data are consistent with the idea that the underlying susceptibility for chromosome nondisjunction involves different components of oogenesis.

#### **2.6.5 Conclusion and Future Directions**

Our candidate gene study was successful in detecting statistically significant associations of maternal nondisjunction of chromosome 21 and variation in genes that are essential for proper chromosome segregation during meiosis. Future studies are needed to investigate other known risk

factors and their interaction with the genetic variation. For example, stratification by maternal age at the time of birth of the infant with trisomy 21 could provide insight into mechanism of the identified genetic variants. In our exploratory analyses, we did not observe unique age-associated variants; however, our sample sizes were limited. Thus, expansion of the study sample with enrichment of the youngest and oldest maternal age groups would be valuable. Also, studies that further stratify meiotic errors by recombination risk patterns known to increase susceptibility of nondisjunction, namely lack of observed recombination, a single telomeric recombination event or a pericentromeric event, may provide further insight into the function of the genetic variant.

Another possible approach is to examine other sources of samples from which information on aneuploidy may be drawn to obtain larger sample sizes. For example McCoy et al. [108] studied day-3 embryos obtained from *in vitro* fertilization cycles and parents to identify both meiotic and mitotic segregation errors. Irrespective, we have begun to gain insight into which meiotic proteins may be more susceptible to genetic variation, leading to abnormal chromosome segregation. Independent studies are needed to replicate findings from our GWAS study to further identify novel susceptibility genes.

## **2.7 ACKNOWLEDGMENTS**

The authors thank Karen Schindler for helpful discussions of Aurora kinases.



### **3.0 ANALYSES STRATIFIED BY MATERNAL AGE AND RECOMBINATION FURTHER CHARACTERIZE GENES ASSOCIATED WITH MATERNAL NONDISJUNCTION OF CHROMOSOME 21**

This chapter has been published in *Prenatal Diagnosis*. The article is reproduced here under license from Wiley with only minor formatting and non-scientific changes. The original article and supporting information are available online at <https://doi.org/10.1002/pd.5919>.

Jonathan M. Chernus<sup>1</sup>, Stephanie L. Sherman<sup>2\*</sup>, Eleanor Feingold<sup>1,3\*</sup>

<sup>1</sup> Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. <sup>2</sup> Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America. <sup>3</sup> Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America. \* EF and SLS are joint senior authors.

### **3.1 CHAPTER OVERVIEW**

**Objective:** In our previous work, we performed the first genome-wide association study to find genetic risk factors for maternal nondisjunction of chromosome 21. The objective of the current work was to perform stratified analyses of the same dataset to further elucidate potential mechanisms of genetic risk factors.

Methods: We focused on loci that were statistically significantly associated with maternal nondisjunction based on this same dataset in our previous study and performed stratified association analyses in seven subgroups defined by age and meiotic recombination profile. In each analysis, we contrasted a different subgroup of mothers with the same set of fathers, the mothers serving as cases (phenotype: meiotic nondisjunction of chromosome 21) and the fathers as controls.

Results: Our stratified analyses identified several genes whose patterns of association are consistent with generalized effects across groups, as well as other genes that are consistent with specific effects in certain groups.

Conclusions: While our results are epidemiological in nature and cannot conclusively prove mechanisms, we identified a number of patterns that are consistent with specific mechanisms. In many cases those mechanisms are strongly supported by available literature on the associated genes.

## **3.2 INTRODUCTION**

At least 5% and possibly as many as 20% of human conceptions are aneuploid as a result of errors in meiosis, and the vast majority of these errors occur in maternal meiosis I, which begins in oocytes during fetal development and is arrested in prophase I for years until resumption in adulthood [1-5, 109]. The decades-long timespan for female meiosis and the different timescales and processes involved in completion of meiosis I and meiosis II make maternal age a critical risk factor for nondisjunction but also make each stage susceptible to different error mechanisms [6, 109]. In addition to maternal age, altered patterns of meiotic recombination are also associated

with nondisjunction [8-14]. As one of the few aneuploid conditions with which humans survive to term, trisomy 21 provides crucial insight into human meiotic nondisjunction [5-7, 20].

In our previous work, we performed the first genome-wide association study to find genetic risk factors for maternal nondisjunction of chromosome 21 [18]. There we contrasted groups of mothers of live-born children with standard trisomy 21 derived from maternal nondisjunction with fathers (a convenient control group). Because earlier research suggested that maternal meiosis I (MI) and meiosis II (MII) nondisjunction of chromosome 21 may have some unique etiological factors and some factors in common, we stratified mothers by meiotic stage of error in order to clarify how genetic risk factors contribute to nondisjunction in each stage. In addition to the genome-wide analyses, we also investigated candidate genes and loci known to be involved with meiotic processes or associated with global recombination counts in humans. We observed a number of interesting associations with loci plausibly involved in maternal meiotic nondisjunction.

In this study we extend our earlier work by performing further stratified analyses of the same dataset in order to dissect the genetic associations we found with MI and MII errors. These analyses will help explore whether the MI/MII classification masks heterogeneity in maternal meiotic nondisjunction and may suggest how genes contribute to particular nondisjunction mechanisms. For example, stratifying by maternal age can clarify whether younger mothers have a higher genetic risk burden than older mothers (or different risk factors altogether). Similarly, within MI and MII errors, different genes may contribute to the recombination profiles on chromosome 21 associated with nondisjunction in each stage. These stratifications may also provide a glimpse into how genes contribute to nonstandard mechanisms such as premature separation of sister chromatids (PSSC) and reverse segregation (RS). Because this further

stratification reduces sample sizes, we performed these stratified analyses in a hypothesis-driven manner, only examining genes and loci that have strong prior support based on the previous work.

### **3.3 METHODS**

#### **3.3.1 Study Participants and Ethics Statement**

##### **3.3.1.1 Participants**

The participants of our study were 749 individuals with non-translocation, maternally-derived trisomy 21 and their 1,437 available biological parents. Self-reported race/ancestry was 72% white, 4% Hispanic, 2% African/African-American or Asian, and 23% other or unknown descent. Participants were recruited in the United States through a birth surveillance system for Down syndrome and later through convenience sampling [7] (see also [18]).

##### **3.3.1.2 Ethics Statement**

At each site an IRB approved the study protocol, consent forms, and data sharing. At Emory University, the data and biological sample repository, the IRB approved sample processing and submission to the Center for Inherited Disease Research for genotyping (Emory School of Medicine IRB number IRB00005100). Study data were uploaded to dbGaP (phs000718) with IRB approval.

### 3.3.2 Genotyping, Imputation, and Population Structure

The genotyping, imputation, and investigation of population structure for this study group have been described in our previous study [18]. Briefly, genotypes were assayed with the Illumina HumanOmniExpressExome-8v1-2 array and initially processed at the Center for Inherited Disease Research. Genotypes for 2,186 unique participants were released to the University of Washington Genetic Coordinating Center for quality assurance/control and imputation. QA/QC for GWAS followed standard procedures described by Laurie et al. [38], and imputation to the 1000 Genomes Project's Phase I reference panel [44] was performed with IMPUTE2 [43]. Trisomic SNP genotypes on chromosome 21 were called from the array data using previously described methods [17]. In addition to these SNP genotypes, STR genotypes from an earlier phase of the study were available on chromosome 21 and were used in a few cases to determine the meiotic stage of trisomy, as described below. Principal components of ancestry were calculated to adjust for population structure; the first three were used as covariates in all regression analyses below. Detailed reports on QA/QC, population structure, and imputation are available through dbGaP (phs000718).

Trisomy (full or mosaic) was confirmed by karyotype in nearly all 749 individuals with Down syndrome, but in a few cases a birth record or parent report was used. Trisomic genotype cluster plots were visually inspected for all 12,982 chromosome 21 SNPs; 3,853 SNPs subjectively regarded as showing poor clustering or showing small minimum cluster sizes were excluded from analysis. As a further quality control step, we compared parental and child SNP genotypes on chromosome 21 to confirm consistency with maternal origin of the trisomy. In 16 instances, the SNP genotypes were not broadly consistent with maternal origin. Most of these cases were traceable to mosaic trisomies or other anomalies identified by karyotyping, but we excluded these

families from downstream analyses, which required broadly reliable genotypes in the children. In 7 instances neither parent was genotyped. In all, 615 parent-child trios and 94 mother-child dyads were available for calling recombination events.

### **3.3.3 Determining Meiotic Stage of Error and Recombination on Chromosome 21**

The primary analyses in this study were centered on mothers who had had a nondisjunction event resulting in live birth of a child with full or mosaic trisomy 21. Subgroups of maternal errors were defined by known risk factors: recombination pattern on the nondisjoined chromosome 21 and maternal age.

To do this, we used parent and offspring SNP and STR genotypes on chromosome 21 to determine stage of meiotic error and to call recombination events on chromosome 21 in each family. Because no “gold standard” procedure exists, we developed new methods for calling recombination events. These methods are fully described in the Appendix.

It is important to note that our classification for MI and MII nondisjunction errors is not precise, although it has been useful to identify distinct risk factors for maternal age, recombination pattern on chromosome 21, and environment [8-14]. The classification is based on whether the variants in the pericentromeric q-arm of chromosome 21 contributed by the parent-of-origin of the nondisjunction error remain heterozygous or are reduced to homozygosity. The former is deemed an “MI error”, as there are representations from both parental homologous chromosomes. The latter defines an “MII error”, as only one parental homolog is represented. However, it is now evident that there are many paths to each of these groups. Whole chromosome nondisjunction clearly leads to an “MI error”, but it is also true that premature (or precocious) separation of sister chromatids (PSSC), either at MI or MII, can appear as an MI or MII error, depending on the

random segregation of chromatids at each stage; reverse segregation (RS, in which sister chromatids from both homologues segregate from each other in MI) can also appear as an MI error, or potentially result in euploidy [110, 111]. Further, there is evidence that MII errors are increased among oocytes affected by MI errors, potentially complicating the ability to define risk factors [112]. Lastly, the U-shaped maternal age curve associated with aneuploidy has been found to be the result of the different age-related patterns associated with the underlying mechanisms [113]. Specifically, PSSC and reverse segregation significantly increase with maternal age, while the gain or loss of a whole chromosome at MI appeared to decrease with the age of the woman. Thus, our simplified classification of “MI errors” and “MII errors” will begin to reveal associated genetic risk factors, but it is only the first step.

### **3.3.4 Association Analyses**

In order to avoid the multiple-testing problems that typically affect subgroup analyses, we focused on loci that were statistically significantly associated with maternal MI or MII nondisjunction in our previous study of the same dataset, with the goal of further dissecting those associations and understanding how they may differ across subgroups. Subgroup differences may suggest models of etiology, although such models would require rigorous verification in future studies, both statistical and mechanistic.

In order to understand how genetic associations with maternal nondisjunction of chromosome 21 vary across relevant risk subgroups of mothers, we performed stratified association analyses in seven subgroups (not all disjoint). In each analysis, we contrasted a different subgroup of mothers with the same set of fathers, the mothers serving as cases

(phenotype: meiotic nondisjunction of chromosome 21) and the fathers as controls. Table 3.1 summarizes the association analyses, abbreviations used to refer to them, and the sample sizes.

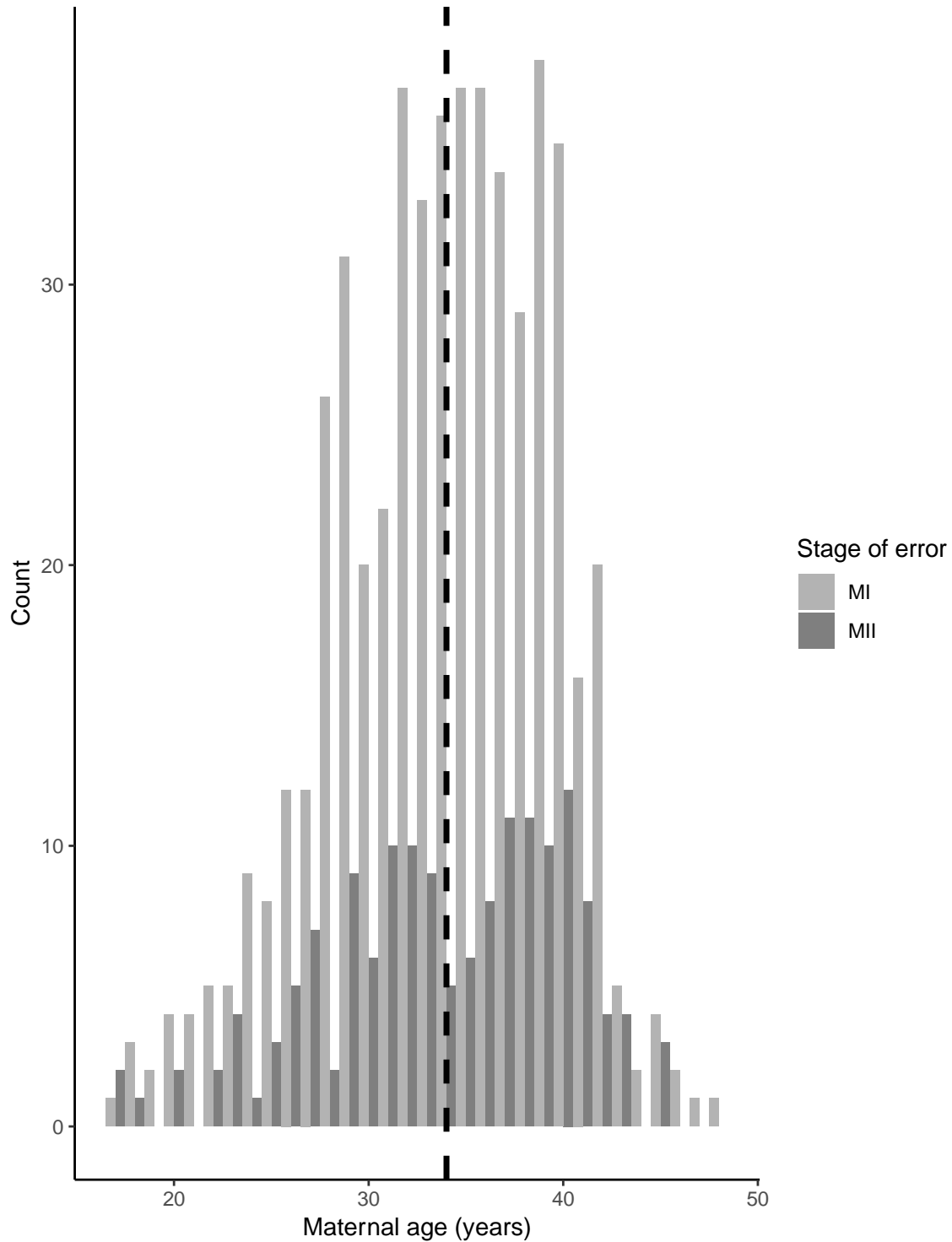
**Table 3.1 Definition of stratified analysis subgroups and associated sample sizes**

Stratification group (fathers are controls)		Abbreviation	Sample size (645 fathers)
Meiotic stage of error	Subgroup		
Meiosis I	0 events	MI R0	328
Meiosis I	Single, telomeric event	MI telomeric	44
Meiosis I	< 34 years old	MI younger	234
Meiosis I	≥ 34 years old	MI older	297
Meiosis II	At least 1 centromeric event	MII centromeric	44
Meiosis II	< 34 years old	MII younger	73
Meiosis II	≥ 34 years old	MII older	82

Recombination events refer here to events observed on chromosome 21.

We first stratified by the meiotic stage in which nondisjunction occurred in the oocyte, MI or MII. We excluded 16 cases in which the error was determined to be an MII with no recombination events detected, as these may represent mitotic rather than meiotic errors. Then, within each error type, we defined subgroups based on maternal age and recombination pattern on chromosome 21. Specifically, mothers with MI and MII errors were each divided into younger and older subsets using a median maternal age split (the median age, 34 years for both MI and MII errors, was included in the older group; see Figure 3.1).



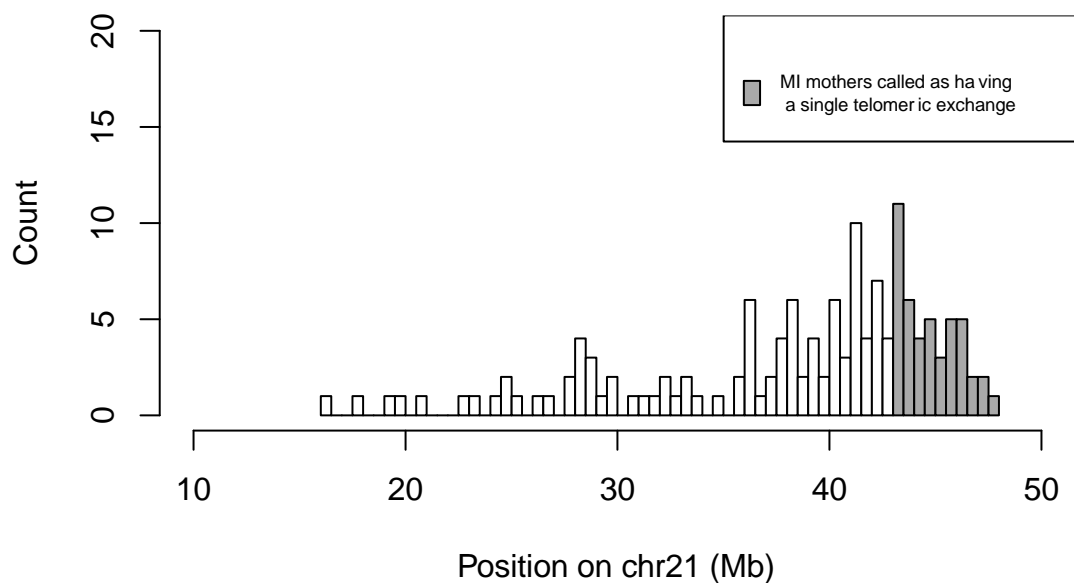


**Figure 3.1 Distribution of maternal ages by meiotic stage of error**

We also defined strata based on the well-established recombination risk profiles associated with nondisjunction: mothers with MI errors with no recombination events observed (“MI R0”),

those with MI errors with a single event near the telomere (“MI telomeric”), and those with MII errors with an event near the centromere (“MII centromeric”). To define telomeric and centromeric events, we plotted the distributions of the positions of single MI events and of the most proximal MII events. Single MI events were defined to be telomeric if they occurred at a position greater than 43Mb. Most-proximal MII events were considered centromeric if they occurred at a position less than 17 Mb (see Figure 3.2).

### Location of single event called in MI mothers



### Location of most proximal event called in MII mothers

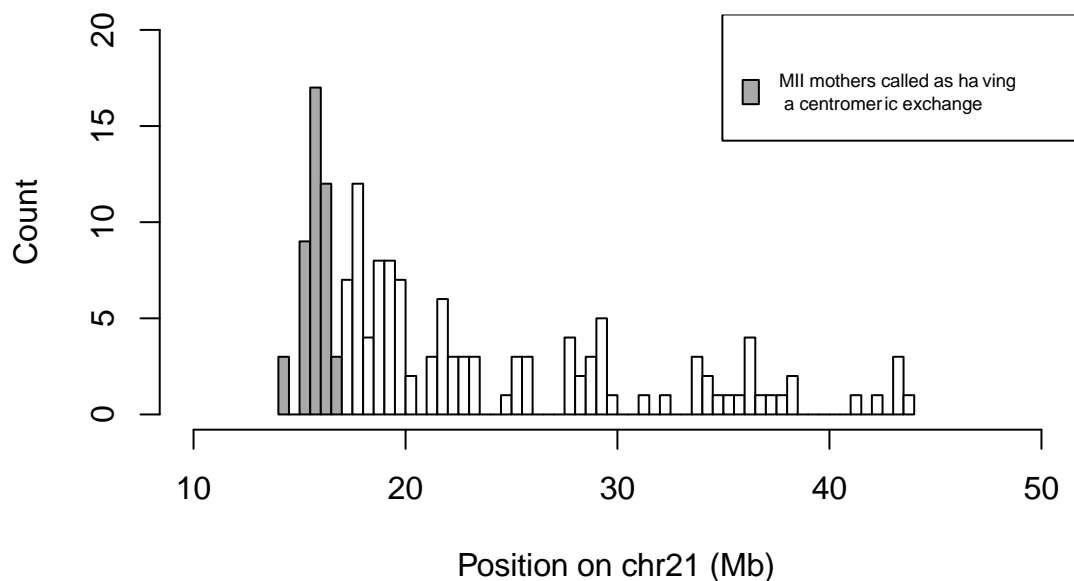


Figure 3.2 Distributions of positions of single MI recombination events and most proximal MII events observed on chromosome 21

For each of these seven subgroups, we tested two sets of loci from our prior analysis of this dataset (discussed in detail in [18]). One set of loci consisted of the 48 individual variants that were identified as associated with nondisjunction in our earlier GWAS of MI and MII nondisjunction. Since our aim in the present study was to explore these pre-recorded associations that had previously attained  $p < 1 \times 10^{-5}$  rather than to identify new loci, we did not apply a multiple-testing correction for the number of single-SNP loci tested. Although these loci were already significant in the MI or MII groups as a whole, we did penalize for the number of subgroups, conservatively applying a significance threshold of  $p = 0.05/7 = 7.14 \times 10^{-3}$  (this is conservative because the subgroups partially overlap). Therefore a subgroup association passing this threshold was considered to be evidence that the subgroup helped to “drive” the earlier GWAS association in the larger group of mothers with MI or MII errors. Note that in these analyses, it is expected to observe nominal significance ( $p < 0.05$ ) in one or more subgroups for each SNP, and our aim is to gain insight by comparing effects across subgroups.

The other set of loci consisted of 37 regions centered around genes relevant to meiotic processes and/or loci associated with global recombination counts in humans [15]. Due to limited sample sizes in the subgroups, we confined this study to these previously-nominated loci instead of performing a new genome-wide scan. Though only a handful of these candidate loci were significant in our previous analysis, we investigated all 37 here in order to uncover possible heterogeneous subgroup effects that may have been masked (with the caveat that weaker new associations must be interpreted cautiously). Since each candidate locus includes numerous SNPs, we used the smallest observed p-value to summarize the association. To adjust for multiple testing at candidate loci, we approximated  $N_{\text{eff}}$ , the effective number of independent SNPs tested at each locus and applied a Bonferroni-corrected threshold for significance ( $0.05/N_{\text{eff}}$ ) [49]. For

conservatively declaring formal statistical significance, we divided this threshold by the number of subgroups (seven) as above.

All association analyses were performed using logistic regression in R [114] and PLINK [115], encoding SNP alleles additively and adjusting for the first 3 principal components of ancestry. In addition to QA/QC filtering described above, SNPs with extreme deviation from Hardy-Weinberg equilibrium and imputed SNPs with info score  $< 0.5$  were excluded. Imputed genotypes called with less than 90% confidence were coded as missing. While all candidate-locus SNPs were subjected to a  $\geq 1\%$  MAF filter (calculated separately for each analysis), the set of 48 GWAS-identified SNPs were not.

We used forest plots to visualize results for single-SNP analyses and LocusZoom [116] plots for candidate regions. We also created two “heatmaps” (Figures 3-4) in order to summarize the results across all GWAS and candidate loci.

### **3.4 RESULTS**

Our previous study of this dataset performed genome-wide and candidate locus association analyses of maternally-derived chromosome 21 nondisjunction. We designed those studies to identify and distinguish variants associated with nondisjunction in maternal MI errors, MII errors, or all maternal errors. Thus, we performed four tests for candidate genes and for the genome-wide scans: mothers with MI errors vs. fathers, mothers with MII errors vs. fathers, all maternal errors combined vs. fathers, and, lastly, mothers with MI errors vs. those with MII errors. Here we refined our analyses into subgroups to further characterize 37 candidate locus associations and 48 GWAS associations resulting from that study. Figures 3.3 and 3.6 summarize the results of our stratified

analyses for the candidate and GWAS loci, respectively. Table 3.2 provides a qualitative summary of selected results. More detailed results are included in the Supporting Information. Below we describe each set of loci in turn.

**Table 3.2 Summary of selected results**

			Subgroups						
			MI R0	MI telomeric	MI younger	MI older	MII centromeric	MII younger	MII older
	Locus	Notes and references							
<b>Prior maternal meiotic nondisjunction GWAS variants</b>	rs10948100 (near <i>VEGFA</i> )	Involved in ovary/follicle development [65, 66]	+	+	+	+	+	+	+
	rs12947774 (near <i>AKAP1</i> )	Involved in meiotic maturation of porcine oocytes (meiotic resumption and MII development) [117]	-	-	-	-	-	-	-
	rs1612273 (near <i>DLGAP2</i> , <i>TDRP</i> )	Variants in <i>DLPGA2</i> are associated with meiotic recombination outside of hotspots [75]; <i>TDRP</i> may be involved in spermatogenesis [76]	+	+	+	+	+	+	+
	rs1855111 (near <i>CXCL12</i> )	Involved in oocyte development in sheep [118]	+		+	+	+	+	+
	rs12652455 (near <i>RICTOR</i> )	Involved in folliculogenesis in mice [119]	+		+	+		+	+
	rs117746305 (in <i>FZD3</i> )	May be involved in follicle growth and oocyte maturation in mice [120]	+		+	+		+	+
	rs148846406 (near <i>EPB41L3</i> )	Possibly involved in folliculogenesis in mice [121]	-	-	-	-		-	
	rs200216460 (in <i>ARMC4</i> )	<i>circARMC4</i> is involved in porcine oocyte meiotic maturation [122]		+		+	+	+	+
	rs11535058 (near <i>SLC39A10</i> )	Also known as <i>ZIP10</i> ; involved in oocyte zinc regulation and meiotic maturation in mice [123]	+			+		+	+
	rs35288347 (near <i>AURKC</i> )	involved in chromosome alignment and other meiotic processes [124]	+		+	+			
	rs9984132 (near <i>COL6A2</i> , <i>PCNT</i> )	Pericentrin is involved in mammalian somatic cell centrosome assembly and meiotic spindle formation in mouse oocytes (with knockdown leading to meiotic spindle disruption, PSSC, and aneuploidy) [73] though it is not present in human meiotic oocytes [74]; Col6a2 is present in the extracellular matrix of cumulus cells in mice [72]	-		-	-			
	rs4649043 (near <i>RUNX3</i> )	Involved in folliculogenesis in mice [125]	-		-	-			
	rs2467011 (near <i>ARID2</i> )	Also known as <i>BAF200</i> ; may be involved in epigenetically reprogramming oocytes; promotes homologous recombination repair of double strand breaks; a specific role in meiosis has not been proven [126, 127]	+		+	+			
	rs16847735 (near <i>CEP70</i> )	Centrosomal protein involved in mitotic spindle assembly [128] (about 500kb from signal)	+		+	+		+	
	rs984968 (near <i>MACROD2</i> )	Haploinsufficiency/knockout causes gross aneuploidies in mouse embryonic fibroblasts and various mitotic segregation defects in human colon cancer cells [129]	+		+	+		+	
rs2560850 (in <i>MYO10</i> )	Involved in mitotic and meiotic spindle assembly [79, 130-133]				+	+	+	+	

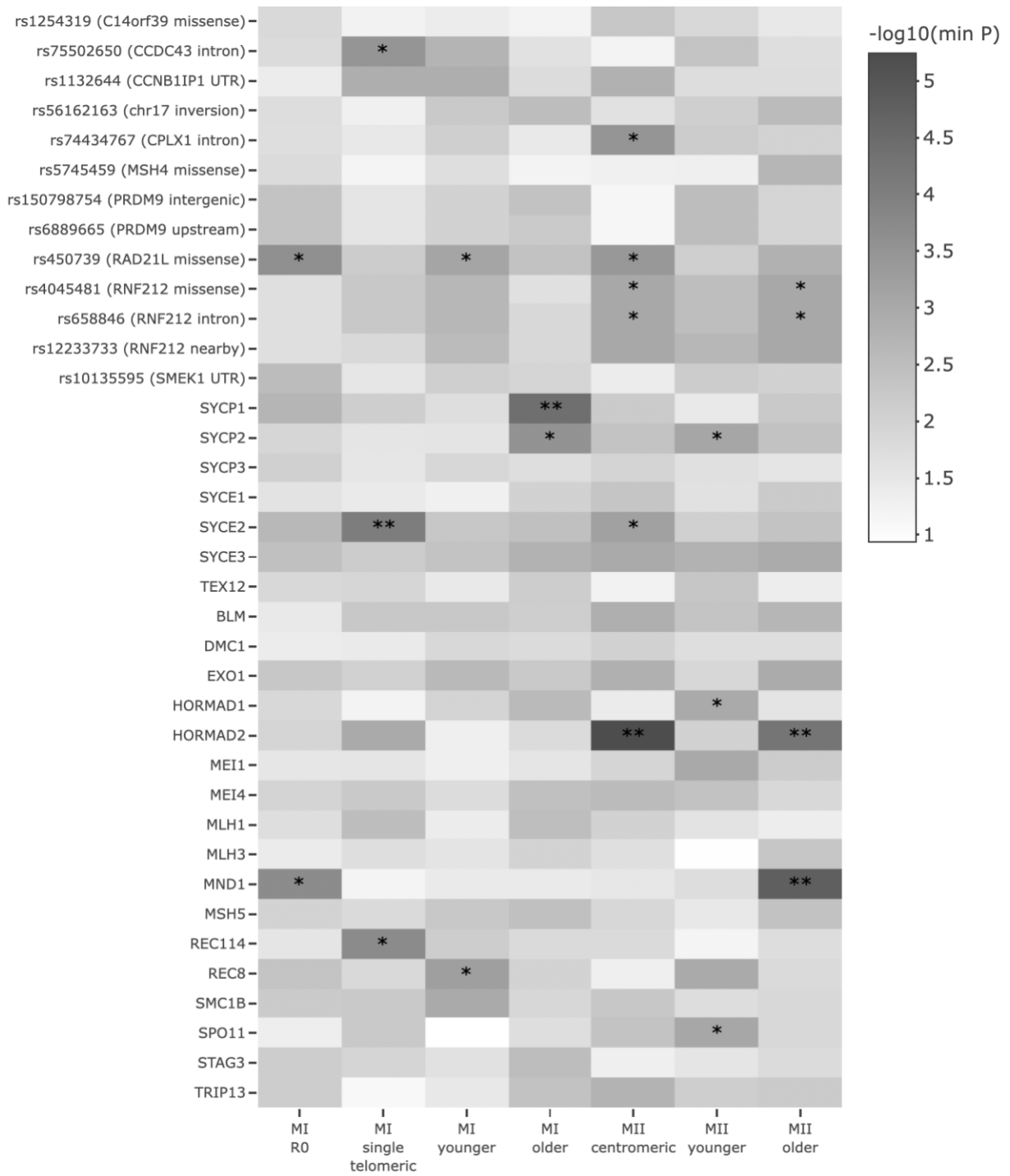
	rs73178888 (near <i>DLGAP2</i> , <i>TDRP</i> )	Variants in <i>DLPGA2</i> are associated with meiotic recombination outside of hotspots [75]; <i>TDRP</i> may be involved in spermatogenesis [76]					+	+	+
	rs12981234 (near <i>XAB2</i> )	Involved in mitotic cell cycle regulation (with knockdown leading to errors in chromosome alignment and segregation) and end resection in homologous recombination [134, 135]					+	+	+
	rs62359711 (near <i>DAB2</i> )	May be involved with porcine oocyte maturation (particularly MII) [136, 137]					+	+	+
	rs115281615 (near <i>CPEB2</i> )	Involved in meiotic maturation of porcine oocytes [138]						+	+
	rs9966603 (near <i>HAUS1/HEI-C</i> )	Involved in mitotic division [139]					+	+	+
<b>Candidate regions</b>	<i>RAD21L</i> region (around rs450739)	Associated with genome-wide recombination counts [15]; meiotic cohesin subunit [50-53, 55]	*		*		*		
	<i>MND1</i>	Involved in processing meiotic double strand breaks and recombination [29, 57-60]	*						**
	<i>SYCE2</i>	Component of central element of synaptonemal complex		**			*		
	<i>SYCP2</i>	Lateral element component of synaptonemal complex				*		*	
	<i>CCDC4</i> region (around rs75502650)	Associated with genome-wide recombination counts [15]; near <i>FZD2</i> , which may be involved in follicle growth, oocyte maturation in mice [140, 141]		*					
	<i>REC114</i>	Influences meiotic recombination by involvement in double-strand break location (in a manner dependent on distance to the telomere on short chromosomes in yeast) [28, 142]		*					
	<i>REC8</i>	Meiotic cohesin subunit [54, 55]			*				
	<i>SYCP1</i>	Transverse filament component of synaptonemal complex				**			
	<i>HORMAD2</i>	Involved in chromosome synapsis/meiotic recombination [26]					**		**
	<i>RNF212</i> region (around rs4045481, rs658846)	Associated with genome-wide recombination counts [15, 75]; also involved with oocyte quality control [143]					*		*
	<i>CPLX</i> region (around rs74434767)	Associated with genome-wide recombination counts [15]					*		
	<i>HORMAD1</i>	Involved in chromosome synapsis/meiotic recombination [26, 27]						*	
	<i>SPO11</i>	Involved in meiotic double strand break formation [28]						*	

Large + and - denote formally significant associations for prior GWAS variants ( $p < 0.05/7 = 7.14 \times 10^{-3}$ ), with the effect allele conferring risk or protection for nondisjunction, respectively. Small + and - denote nominal significance ( $p < 0.05$ ). \*\* Denotes significant associations for candidate loci (correcting for the number of independent variants tested and the number of subgroups). \* Denotes nominal significance for candidate regions (correcting for the number of independent variants only).

### 3.4.1 Candidate Loci

The 37 candidate loci we tested comprise two groups. The first group consists of 13 regions surrounding SNPs found by Kong et al. to be associated with global recombination counts in a large Icelandic population [15]. The second group consists of 24 candidate genes involved with recombination and other meiotic processes (see [18] for a discussion of these). The “heatmap” in Figure 3.3 summarizes the results across all candidate loci. Of these, 22 associations across 14 loci were nominally significant ( $p < 0.05/\text{number of SNPs at locus}$ ), though only 5 associations across 4 loci passed the conservative threshold. Figures 3.4 and 3.5 show LocusZoom plots for some of these loci, noting that these loci represent small windows around specific SNPs and genes, not single-SNP loci. Additional LocusZoom plots are included in the Supporting Information.





**Figure 3.3 Summary of stratified analysis results for 37 candidate loci**

### 3.4.1.1 Results for Candidate Loci Associated with Genome-Wide Recombination Counts

Among the 13 recombination-associated loci tested in our previous analysis, we found three that were statistically significantly associated with maternal MI or MII nondisjunction: *RAD21L*, *CCDC43* and *SMEK1*. In these stratified analyses, none of those loci showed subgroup associations that passed the conservative significance threshold. In order to describe which subgroups may have driven the previous significant associations, we simply identify the strongest subgroup associations (formally significant or not). The *RAD21L* (rs450739) locus was previously associated with all maternal cases and MI cases. The results of the MI subgroup analyses, while not passing the conservative significance threshold, suggest that the MI R0 and MI younger subgroups may have primarily driven the previous association. The *CCDC43* (rs75502650) locus was previously associated with both MI and MII maternal nondisjunction. Among the MI subgroups, the single telomeric exchange subgroup shows the strongest effect; among the MII subgroups, none is even nominally significant. Though *CCDC43* itself has not otherwise been linked to meiotic processes, the nearby gene *FZD2* may be involved in ovarian follicle/oocyte development [140, 141]. The *SMEK1* (rs10135595) region was the only candidate locus of this group that showed a significant association in our previous GWAS (in the MI vs. MII contrast) but showed only weak associations in the seven subgroups in the stratified analyses.

We observed nominally significant associations for two candidate loci not found significant in our previous analyses. The *CPLX1* (rs74434767) locus was nominally significant in the MII centromeric subgroup. The *RNF212* locus (specifically, windows surrounding rs4045481 and rs658846, located in that gene) was nominally significantly associated in the MII centromeric and MII older subgroups. Regional association plots for the *CCDC43*, *RAD21L*, and *CPLX1/RNF212* loci are shown in Figure 3.4.

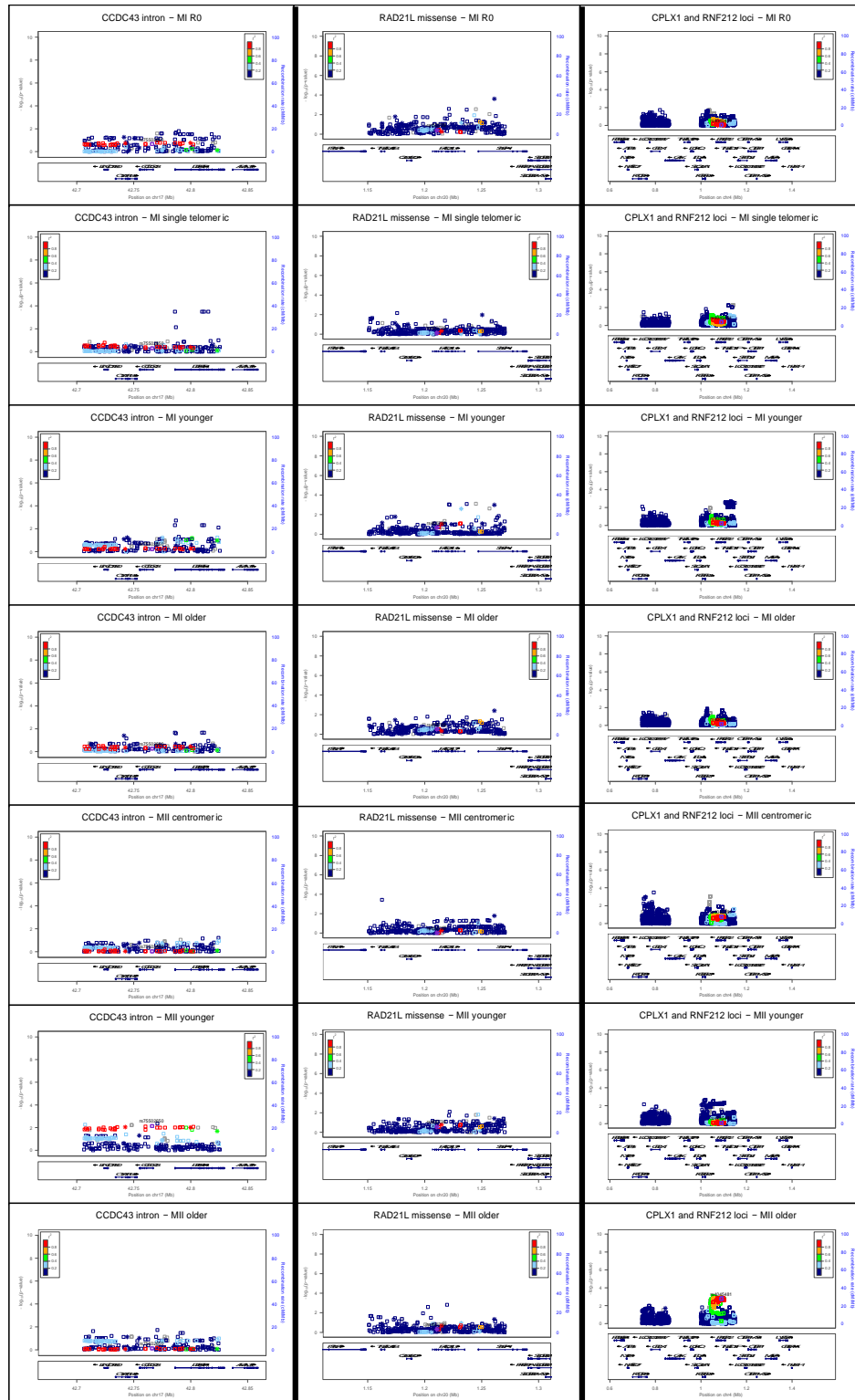


Figure 3.4 LocusZoom plots for selected associations in candidate loci for global recombination (*CCDC43*, *RAD21L*, and *CPLX1/RNF212* loci)

### 3.4.1.2 Results for Candidate Loci Involved with Meiotic Processes

Among the 24 meiosis/recombination gene loci tested for maternal nondisjunction in our previous work, four were significant: *SYCP1* (MI vs. MII), *SYCP2* (MII and MI vs. MII), *SYCE2* (MI), and *MND1* (MII and MI vs. MII). In our stratified analyses, we found significant associations in three of these four genes and in an additional gene not previously found associated with maternal nondisjunction (*HORMAD2*). The stratified analyses honed in on associations with specific subgroups in genes that play a role in the synaptonemal complex: *SYCP1* (MI older), *SYCP2* (the strongest associations, only nominally significant, were in the MI older and MII younger subgroups), and *SYCE2* (significant for MI single telomeric and nominally for MII centromeric). *HORMAD2*, which plays a role in coordinating progression of chromosome synapsis with meiotic recombination, was significant in the MII centromeric and MII older subgroups. Regional association plots for *SYCP1*, *SYCP2*, and *MND1* are shown in Figure 3.5, and the remaining loci are shown in the supporting information.

There were several nominally significant associations in genes that were not previously significantly associated with maternal nondisjunction. Some of these associations involved genes with a role in DNA double strand breaks, the precursor to recombination. These include *SPO11*, that was nominally significant in the MII younger subgroup, *REC114* that was nominally significant in the MI single telomeric group and *MND1* that was nominally significant in the MI R0 and MII older subgroups. Lastly, *REC8*, a component of the meiotic cohesin complex, was nominally significantly associated in the MI younger subgroup.

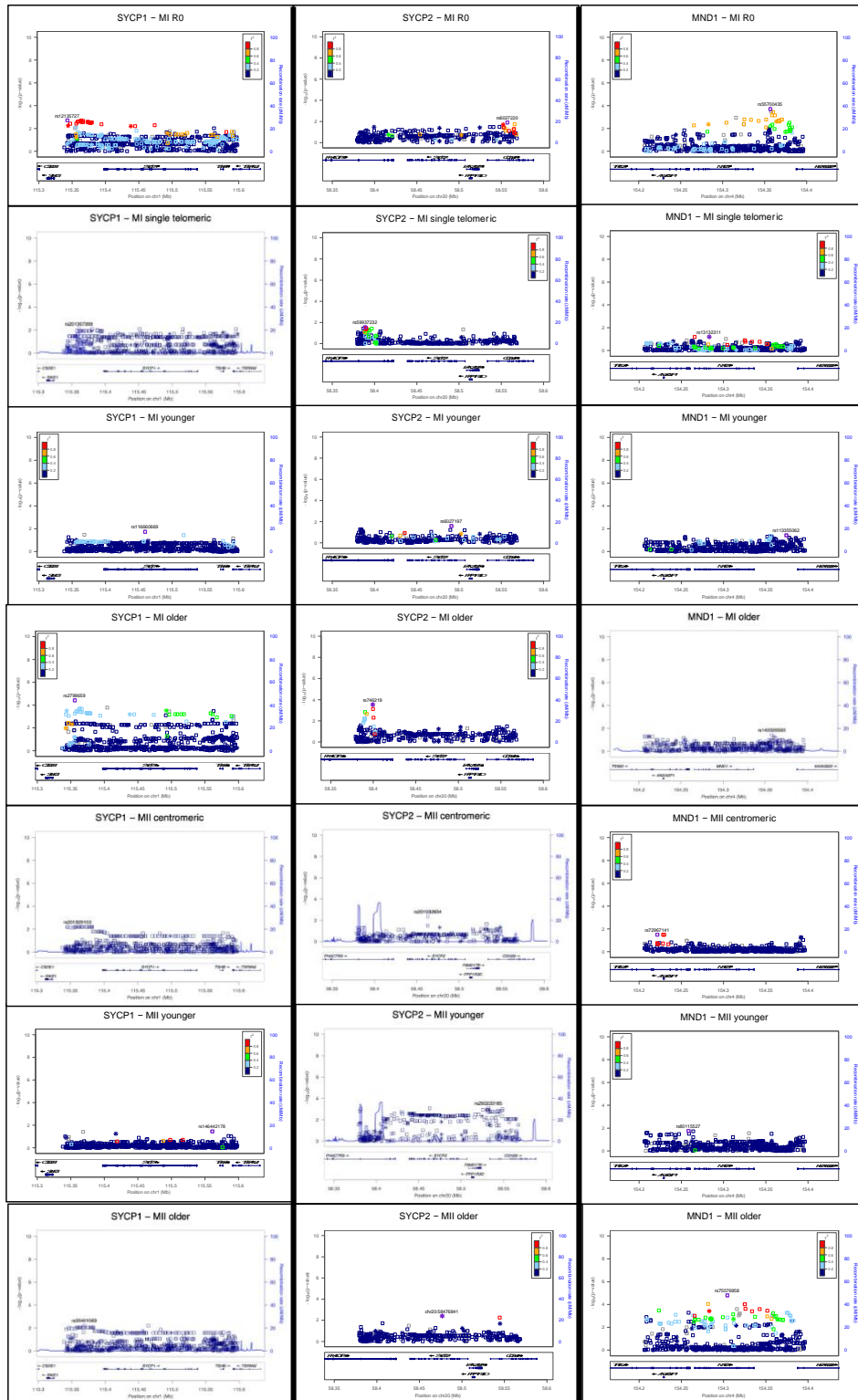
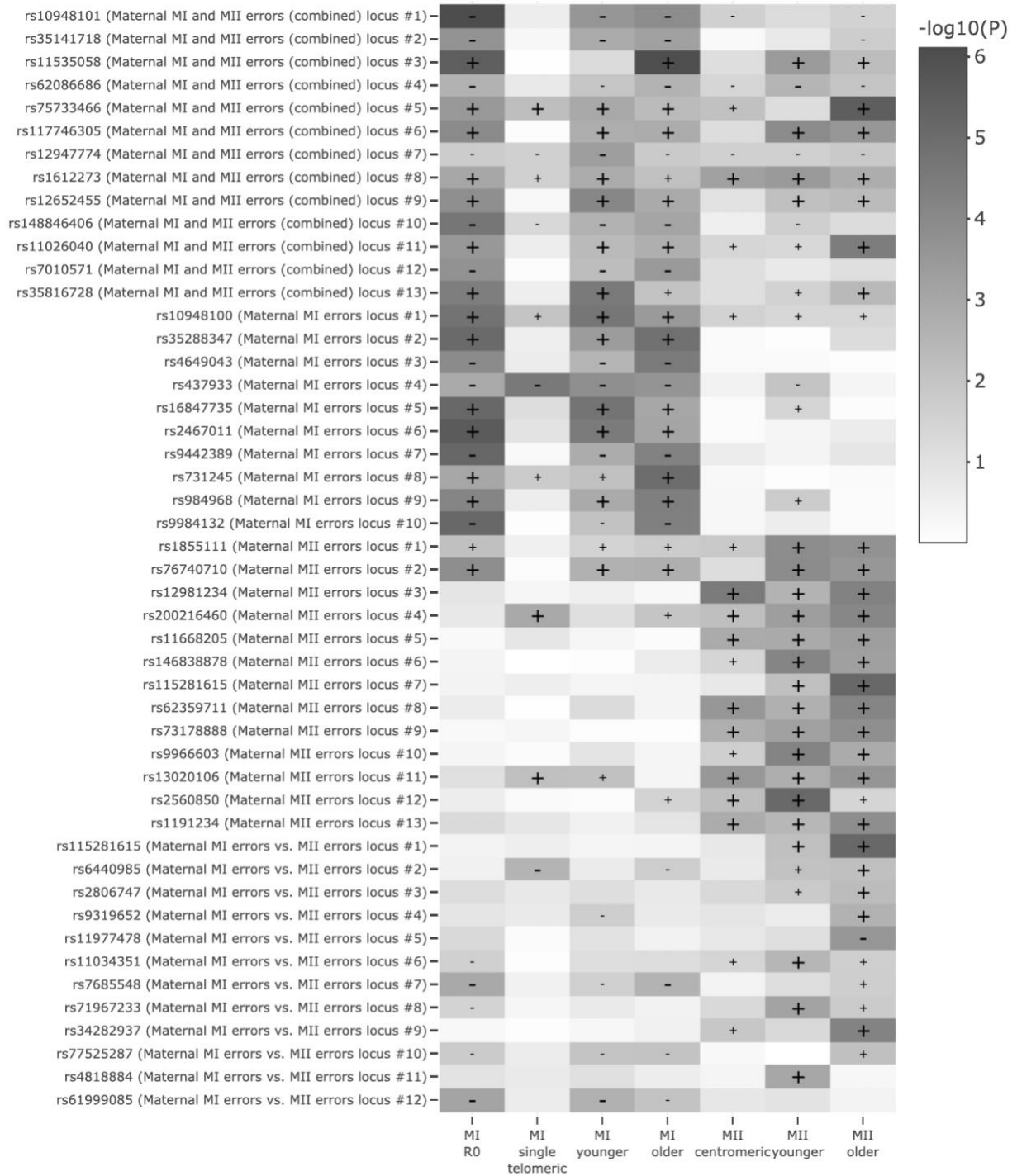


Figure 3.5 LocusZoom plots for selected associations in candidate genes involved in meiotic processes (*SYCP1*, *SYCP2*, and *MND1*)

### 3.4.2 Prior chromosome 21 nondisjunction GWAS associations

Based on our previous study design using four analyses to understand MI- and MII-specific risk factors, 48 SNPs were nominated as associated with nondisjunction in MI, MII, or both. In the present study we performed stratified analyses of these SNPs. For each prior association, we investigated whether the strength and direction of that SNP's association varied across the seven strata, as summarized in Figure 3.6. Examples of forest plots summarizing the associations at four selected loci are shown in Figure 3.7 (note that for three of the four loci shown, the effect allele generally confers risk; for the *VEGFA* locus, the effect allele appears protective). Below we discuss the results in detail, grouping the SNPs according to their prior associations. The Supporting Information includes a table of results for each subgroup as well as a forest plot for each GWAS locus.



**Figure 3.6 Summary of stratified analysis results for 48 previous nondisjunction GWAS loci**

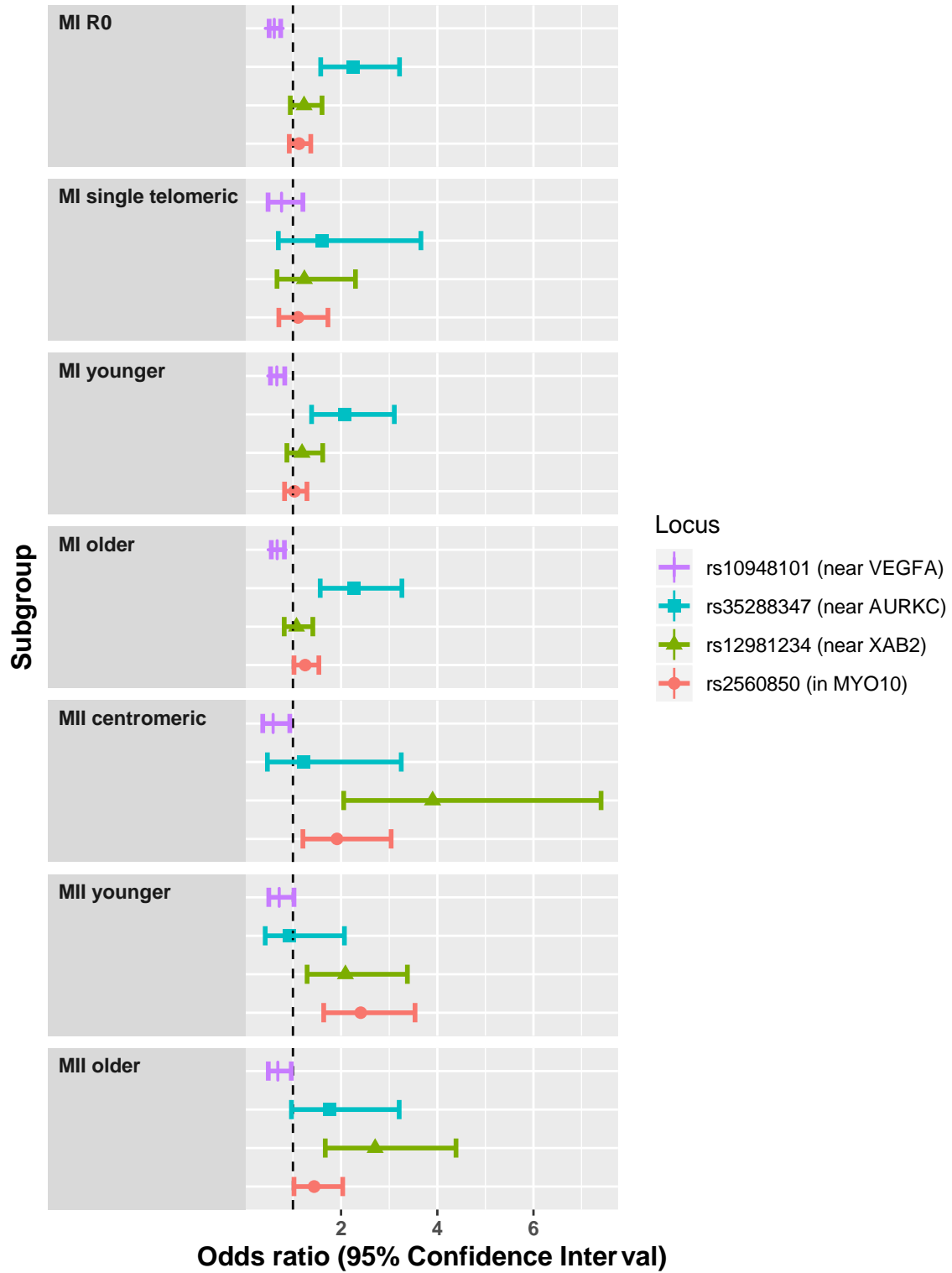


Figure 3.7 Forest plots for four selected prior GWAS loci



### 3.4.2.1 Results for Top Associations from Previous All Maternal (MI and MII) Errors

#### Combined Analysis

Among the 13 variants previously identified from the combined MI and MII mothers vs. fathers analysis, we observed consistent estimated directions of effect for all seven subgroups (i.e., for each SNP, either all odds ratios were estimated to be  $< 1$  or all were estimated to be  $> 1$ , though confidence interval widths varied). This is unsurprising, given that these variants were identified because they were associated with the phenotype of maternal nondisjunction, combining both MI and MII errors. However, it is also possible that strong effects in one or more subgroup “drive” these associations, with weak or null effects in others. Here we saw that the MI R0, MI younger, and MI older subgroups (three of the larger subgroups) passed the conservative significance threshold ( $p < 7.14 \times 10^{-3}$ ) for 12, 11, and 10 of the 13 variants, respectively. The two smallest subgroups (MI single telomeric and MII centromeric) each passed the conservative threshold for only one subgroup. Given the lack of heterogeneity in the point estimates, the pattern of significance we observe across subgroups appears broadly consistent with homogeneous effects with low power in smaller subgroups (though our analysis is not designed to prove this).

Here we highlight some associated SNPs that occur near genes that play roles in follicle or oocyte development. rs11535058 near *SLC39A10* (also known as ZIP10, involved in the oocyte-to-egg transition [123]) was significantly associated in the MI R0 and older subgroups as well as the MII younger and older subgroups (Figure 3.6, Maternal MI and MII (combined), locus #3). rs12652455 near *RICTOR* (involved in folliculogenesis in mice [119]) was significantly associated in all but the MI telomeric and MII centromeric subgroups (Figure 3.6, Maternal MI and MII (combined), locus #9). rs148846406 near *EPB41L3* (Figure 3.6, Maternal MI and MII (combined),

locus #10) was significantly associated in the three largest subgroups (MI R0, MI younger, MII older). *EPB41L3* may have a role in folliculogenesis [121].

No loci passed the conservative threshold for across all seven subgroups, though two attained nominal significance ( $p < 0.05$ ) in all subgroups: rs12947774 (near *AKAP1*, which has a role in porcine oocyte maturation [117]; Figure 3.6, Maternal MI and MII (combined), locus #7) and rs1612273 (near *DLGAP2*, found to be associated with recombination outside of hotspots [75]; also near *TDRP*, possibly involved in spermatogenesis [76]; Figure 3.6, Maternal MI and MII (combined), locus #8). Another variant at this locus is discussed below as an MII association.

In our previous analyses, several of the strongest associations from the combined MI and MII analysis also showed evidence of an effect in the stratified MI and MII error analyses, sometimes at a nearby variant rather than at the same variant. The most significant association from our previous GWAS was rs10948101 (Figure 3.6, Maternal MI and MII (combined), locus #1), near *VEGFA* (involved in folliculogenesis [65, 66]), which was significant in the three largest subgroups (MI R0, MI younger, and MI older). Another SNP nearby (rs10948100, Figure 3.6, Maternal MI errors locus #1) showed the same pattern. Similarly, rs117746305 (Figure 3.6, Maternal MI and MII (combined), locus #6) in *FZD3* was near rs76740710, an MII-associated locus (Figure 3.6, Maternal MII errors locus #2). Recent research has shown that Frizzled 3 may have a role in follicular development and oocyte maturation [120]. These two *FZD3* variants were significantly associated with nondisjunction in all but the MI telomeric and MII centromeric subgroups.

### **3.4.2.2 Results for Top Associations from Previous Maternal MI Errors Analysis**

Among the 10 MI-specific prior associations, there appears to be little evidence of different effects across the MI R0, MI younger, or MI older groups. Only in one instance does the direction

of a point estimate differ among subgroups. rs35288347 near *AURKC* (involved in chromosome alignment and other meiotic processes [124]; Figure 3.6, Maternal MI errors locus #2) shows significance in all MI subgroups except the single telomeric exchange subgroup. rs4649043 and rs2467011 (Figure 3.6, Maternal MI errors locus #3 and #6, near *RUNX3* and *ARID2*, respectively) follow the same pattern. *RUNX3* is involved in folliculogenesis in mice [125]. *ARID2* (also known as BAF200) may be involved in epigenetically reprogramming oocytes and has been shown to promote homologous recombination repair of double strand breaks, though a specific role in meiosis has not been proven [126, 127]. rs9984132, similarly, passes the conservative significance threshold in the MI R0 and MI older subgroups, but attains only nominal significance in the MI younger subgroup (Figure 3.6, Maternal MI errors loci #10). This association is near *PCNT*, which encodes pericentrin, a protein involved in mammalian somatic cell centrosome assembly and meiotic spindle formation in mouse oocytes, with knockdown leading to meiotic spindle disruption, PSSC, and aneuploidy [73].

The smallest subgroup, MI single telomeric, passes the conservative significance threshold in only one case, generally tracking the other subgroups but with a wider confidence interval. This significant association occurs at rs437933 (Figure 3.6, Maternal MI errors locus #4), discussed further below. Though none of the MII subgroups show a significant effect ( $p < 7.14 \times 10^{-3}$ ) at any of these MI loci, the MII younger subgroup confidence intervals for three MI loci may suggest a (possibly weaker) effect, tracking the significant effects in the MI R0, MI younger, and MI older subgroups, (rs437933, rs16847735, and rs984968; Figure 3.6, Maternal MI errors locus #4, #5, and #9). Of these, rs437933 (mentioned above), which has the strongest MI single telomeric association among all 48 loci, is located near *THEG5* (“testis highly expressed protein 5”), which is poorly characterized; rs16847735 is located about 500kb from *CEP70* (a centrosomal protein

involved in mitotic spindle assembly [128]); and rs984968 is in *MACROD2* (haploinsufficiency/knockout of which causes gross aneuploidies in mouse embryonic fibroblasts and various mitotic segregation defects in human colon cancer cells [129]).

### 3.4.2.3 Results for Top Associations from Previous Maternal MII Errors Analysis

Of 13 variants associated with maternal MII nondisjunction in our previous analyses, eight show generally consistent effects across all MII subgroups and little evidence of effect in any MI groups. For all 13 loci, effect directions are the same across all MII subgroups, though the MII centromeric group generally has larger confidence intervals and lower significance, reflecting its smaller sample size. For instance, rs115281615 near *CPEB2* shows consistent MII subgroup associations with no significant associations in any MI subgroups (Figure 3.6, Maternal MII errors locus #7). This gene is necessary for meiotic maturation of porcine oocytes, particularly for reaching MII metaphase [138]. rs62359711 near *DAB2* (Figure 3.6, Maternal MII errors locus #8), which may be involved in oocyte maturation, particularly at MII [136, 137], follows a similar pattern. rs73178888 (Figure 3.6, Maternal MII errors locus #9), also significantly associated with nondisjunction in all three MII subgroups, is located near *DLGAP2* but is not in linkage disequilibrium with rs1612273 (discussed above in the combined MI and MII analysis); therefore, these associations may represent independent signals. rs9966603 (Figure 3.6, Maternal MII errors locus #10) is located near *HAUS1* (also known as *HEI-C*), which is required for proper mitotic division, with depletion resulting in an abnormal mitotic spindle [139]. The strongest association for the MII centromeric group was at rs12981234 (Figure 3.6, Maternal MII errors locus #3). This SNP lies in a gene-rich region that includes *XAB2*. *XAB2* has numerous functions, including mitotic cell cycle regulation (with knockdown leading to errors in chromosome alignment and segregation) as well as end resection in homologous recombination [134, 135].

Several of the remaining loci may suggest shared effects in some MI subgroups, with the added caveat that some of these MI associations are only nominally significant, without passing the conservative threshold. These include rs76740710 in *FZD3*, mentioned above (Figure 3.6, Maternal MII errors locus #2), significant in the MI R0, MI younger, and MI older subgroups. Another example is rs200216460 in *ARMC4*, which is significantly associated in all MII subgroups as well as the MI telomeric group, with nominal significance in the MI older group (Figure 3.6, Maternal MII errors locus #4). Recent work has shown that *ARMC4* circular RNAs are crucial to oocyte maturation in pigs [122]. The top MII variant, rs1855111 (in *TMEM72-AS1*; Figure 3.6, Maternal MII errors locus #1), attained nominal significance in the MI R0, MI younger, and MI older groups, but did not pass the conservative threshold. Although neither *TMEM72-AS1* nor the nearby *TMEM72* appear to be involved in meiosis, the gene *CXCL12* (470kb downstream) may play a role in ovine oocyte development [118]. Finally, rs2560850 in *MYO10* is significantly associated in the MII centromeric and MII younger subgroups and attains nominal significance in the MI older and MII older subgroups (Figure 3.6, Maternal MII errors locus #12). In addition to roles in mitotic spindle function and mitotic progression, *MYO10* is required for meiotic spindle assembly [79, 130-133].

#### **3.4.2.4 Results for Top Associations from Previous Maternal MI Errors vs. MII Errors**

##### **Analysis**

For the variants we found associated from the “MI vs. MII” previous analyses, we expected to observe evidence of effects in the opposite direction among the associations in MI and the MII subgroups. Among the 12 variants analyzed, six followed this pattern: they were at least nominally significantly associated in at least one MI subgroup and at least one MII subgroup, with opposite directions of effect. Of these six, four suggested (at least) a contrast between the MI R0 group and

the MII older group (Figure 3.6, Maternal MI errors vs MII errors locus #6, #7, #8 and #10). Another suggested opposite significant effects for the MI younger and MII older subgroups (Figure 3.6, Maternal MI errors vs MII errors locus #4); the last of these contrasted the MI telomeric and older groups with the MII younger and older groups (Figure 3.6, Maternal MI errors vs MII errors locus #2). The remaining six variants did not show the expected pattern for an “MI vs. MII” variant, although there were noted associations within MI error (Figure 3.6, Maternal MI errors vs MII errors locus #12) or MII error (Figure 3.6, Maternal MI errors vs MII errors locus #1, #3, #5, #9 and #11) subgroups.

### **3.5 DISCUSSION**

The purpose of this study was to better understand maternal meiotic disjunction of chromosome 21 by following up our previous findings from a GWAS and candidate gene study. In that study, we stratified maternal nondisjunction based on MI- or MII-derived errors (see Introduction and below). Here, we further examined heterogeneity within those groups based on known risk factors, namely maternal age and recombination profile on chromosome 21 in order to gain further insight into potential modes of action for each gene. While our study was not designed to prove or disprove any particular mechanism, it does complement recent important research by investigating how common genetic variation may contribute to these mechanisms. Below we discuss our results within the context of the proposed mechanisms and the limitations of our approach.

The simplest conceptual category of genetic risk factors is those that cause general nondisjunction risk. Such variants would broadly increase the risk of an error across all subgroups

of MI and MII nondisjunction, irrespective of risk factors such as age and recombination. The prior GWAS associations we followed up on here are enriched for variants near oocyte development or folliculogenesis genes (Table 2), and the results of our stratified analyses indicate that such genes are consistent with this model of contributing to meiotic nondisjunction errors broadly. For example, the *VEGFA*, *AKAP1*, and the *DLGAP2* (rs1612273) loci all have at least nominally significant associations with the same effect directions across all subgroups. Additionally, the *SLC39A10*, *RICTOR*, *FZD3*, *CXCL12*, *ARMC4*, and *EPB41L3* loci may follow a similar pattern, but with small sample sizes limiting power in some subgroups. Conversely, it should be noted that consistent associations across subgroups for a given locus do not necessarily imply the same mechanism across subgroups, since such results are also consistent with pleiotropic effects.

We also observed results that are consistent with standard MI and MII errors and their associated recombination profiles (for MI, either no observed recombination on chromosome 21, or a single telomeric, exchange; pericentromeric recombination for MII). Both types of errors are associated with advanced maternal age. With respect to the altered recombination patterns, the MI-associated patterns are independent of maternal age while MII-associated pericentromeric recombination increases with increasing maternal age [11, 13]. Our results suggest that genetic variants may contribute to standard MI or MII nondisjunction mechanisms (these could be either the common variants we tested or unobserved variants, possibly rare, in linkage disequilibrium with them). This interpretation is based on functions of the implicated genes and the patterns of associations across subgroups. For example, the *AURKC* locus is broadly associated with nondisjunction in MI (significantly in the younger, older, and R0 subgroups). While the functions of *AURKC* in meiosis are complex and not fully understood, loss of *AURKC* function in mouse oocytes can cause misalignment of chromosomes at MI and aneuploidy in MII oocytes [124]. On

the other hand, the *XAB2* and *HAUS1/HEI-C* loci were both broadly associated with nondisjunction in MII subgroups. *XAB2* and *HAUS1* are required for proper mitotic division, which has important similarities with MII division. These results are consistent with a hypothetical mechanism in which defects in machinery shared with mitotic processes cause failure of sister chromatids to segregate at MII.

Additionally, several candidate loci were associated specifically with nondisjunction in MI or MII recombination subgroups, suggesting these loci could contribute to standard MI or MII errors by increasing the likelihood of “susceptible” recombination patterns on chromosome 21. For instance, the previously observed *SYCE2* (a synaptonemal complex element) association appears to be driven primarily by the MI single telomeric and the MII centromeric subgroups. Another possible example is *REC114*, nominally significantly associated only in the MI single telomeric subgroup, which influences meiotic recombination in yeast by controlling double-strand break location in a manner dependent on distance to the telomere on short chromosomes [142]. (The nominal *REC114* association should be interpreted cautiously, since *REC114* was not significantly associated with nondisjunction in our previous study). Contrasting the *SYCE2* and *REC114* associations with the loci discussed above as possible candidates for standard nondisjunction, we observe evidence of potentially important genetic heterogeneity. Whereas *SYCE2* and *REC114* appear to be associated only in recombination-defined subgroups, the *AURKC*, *XAB2*, and *HAUS1/HEI-C* loci are broadly associated across MI or MII subgroups, with significant associations both in subgroups defined by recombination and those not. Together these observations suggest that some variants in/near meiosis genes could contribute “narrowly” to nondisjunction risk by contributing specifically to aberrant recombination, while others may



contribute broadly, independently adding to (or possibly interacting with) recombination-associated risk within a particular stage of meiosis.

Several other results in the recombination-defined subgroups are more difficult to interpret and merit discussion. We note that among the global recombination count loci identified by Kong et al. [15], only the *RAD21L* locus was even nominally associated in the MI R0 subgroup, where we might have expected several other recombination-count genes to be associated, given evidence that lower genome-wide recombination is associated with aneuploidy in oocytes [111]. Another counterintuitive finding is that two previous synaptonemal complex gene associations appeared to be driven primarily by subgroups not defined by recombination profile (specifically, *SYCP1* for the MI older subgroup, and *SYCP2* for the MI older and MII younger subgroups).

Our results give limited insight into issues around PSSC and RS mechanisms. The simplistic classification of meiotic nondisjunction events as “MI errors” or “MII errors” based solely on whether homologous chromosomes vs. sister chromatids are passed from mother to child is an inference from incomplete data, since effectively only one product of female meiosis is observed (the oocyte). Without further information about the chromosomal content of the other products of meiosis (the first and second polar bodies), the sequence of events leading to aneuploidy in the oocyte cannot be directly inferred. Indeed, other mechanisms such as premature separation of sister chromatids (PSSC) or reverse segregation (RS) are major causes of aneuploidy in oocytes but are indistinguishable from standard MI or MII errors in our study. It is important to note that although both mechanisms involve separation of sister chromatids in MI, the relative rates of PSSC and RS suggest that the latter is not simply due to two independent occurrences of the former [111]. This is further corroborated by the more extensive loss of cohesion involved in RS than in PSSC. These observations suggest that the two mechanisms could either share genetic

factors (influencing, e.g., cohesion in general) or have some unique factors (affecting, e.g., cohesion at the centromere vs. chromosome-wide).

These additional mechanisms likely explain some of the heterogeneity in our results. Since PSSC can lead to apparent MI or MII errors, variants affecting the risk of PSSC may be represented in our results as signals shared by MI and MII subgroups. Additionally, since lack of recombination may contribute to PSSC and RS [111], signals shared with the MI R0 subgroup may implicate these mechanisms. This could explain, for instance, why the previous *MNDI* association appears to be driven by the MI R0 and MII older groups. Our candidate gene results also suggest that age-related associations in cohesin genes could represent PSSC and/or RS. For instance, *REC8* is nominally associated in the MI younger group (with a similar but non-significant effect in the MII younger group; note, however, that *REC8* was not significant in our previous analysis). The prior association with the cohesin subunit gene *RAD21L* (also associated with genome-wide recombination [15]) appeared to be driven by associations in the MI R0, MI younger, and MII centromeric groups. These signals suggest that variants at this locus could contribute to multiple mechanisms: increasing the risk of MI nondisjunction due to lack of recombination, promoting sister chromatid separation through age-related degradation of cohesin, or exacerbating pericentromeric recombination.

Among the potentially heterogeneous associations we observed in the GWAS loci were several involving maternal age subgroups. rs437933, rs16847735, and rs984968 (near *THEG5*, *CEP70*, and *MACROD2*, respectively), which are each significantly associated with nondisjunction in the MI R0, MI younger, and the MI older groups, and nominally associated in the MII younger group but with null effects in the MII older group. rs2560850 (near *MYO10*) was significantly associated with nondisjunction in MII subgroups and nominally in the MI older

group, but with no effect in the MI younger group. If genuine, these heterogeneous signals could represent genetic effects contributing to apparent MI and MII errors through age-dependent effects on PSSC. But further work is needed to characterize these complex associations, especially given that some are only nominally significant.

It is important to note that this study had limited ability to detect heterogeneous effects among subgroups precisely because of how the loci were chosen. The 48 maternal meiotic nondisjunction associations we investigated first came to our attention because our prior GWAS that stratified only by type of nondisjunction error: MI or MII. In order to be statistically significant in this previous analysis, it is expected that effects within each error type were in the same direction; otherwise, the heterogeneity would be masked. Thus it is unsurprising that we did not generally observe statistically significant associations with opposite directions of effect for MI and MII error subgroups.

### **3.6 CONCLUSION**

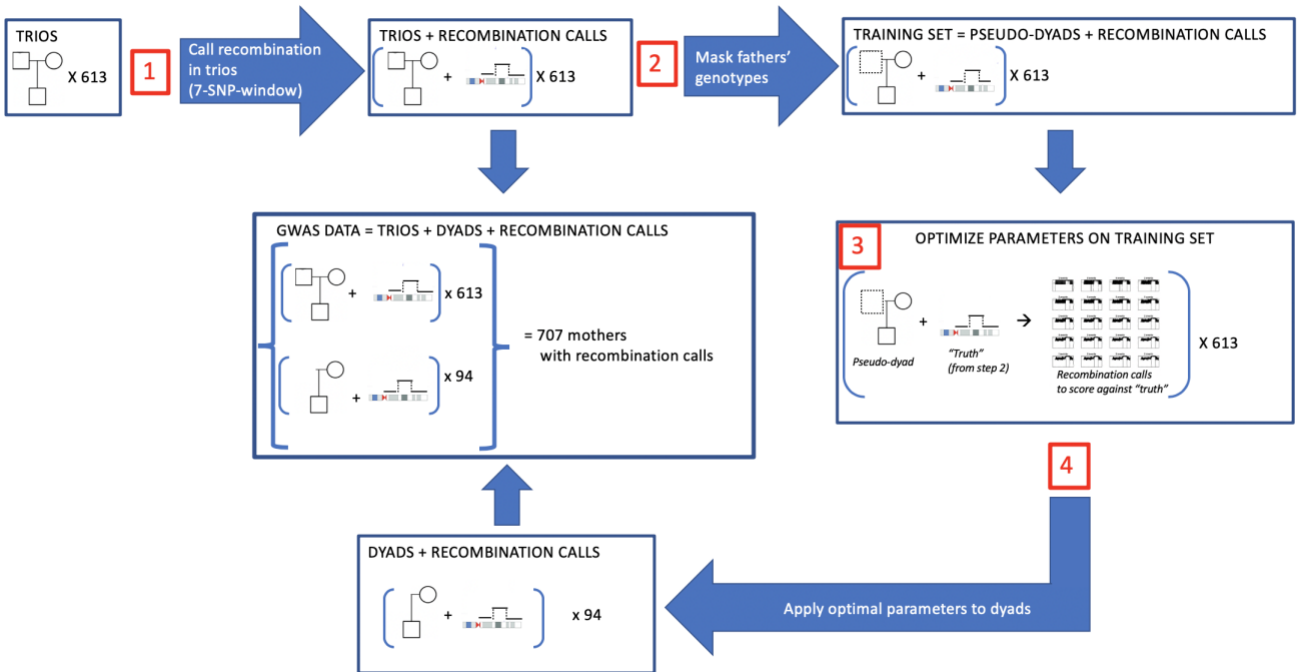
Our stratified analyses succeeded in uncovering heterogeneity in earlier genetic associations. While our approach of classifying errors as MI or MII in origin does not enable us to infer nondisjunction mechanisms in particular cases, our results provide insight into how variation in oocyte development and meiotic processes may contribute to specific nondisjunction mechanisms. In future studies, larger sample sizes will enable efforts to replicate the reported associations and perhaps even genome-wide scans in the smallest subgroups. In the future it may also be possible to perform a similar study in which all products of female meiosis are recovered, allowing inference of the nondisjunction mechanism in each case and more refined stratifications.

Finally, we are planning analyses of altered maternal recombination on chromosome 21 as an outcome in itself, which may have higher power for detecting genetic variants contributing to nondisjunction risk and may therefore clarify our understanding of how recombination affects that risk.

## **3.7 APPENDIX: METHODS FOR CALLING MEIOTIC STAGE OF NONDISJUNCTION AND RECOMBINATION EVENTS**

### **3.7.1 Overview**

We used a different method for calling stage of origin and recombination events in trios (with child and both parents genotyped) than in dyads (with mother and child genotyped, but not the father), since the latter contain less information. Because our analyses focus on recombination in oocytes, we did not call recombination events in father-child dyads. Our overall approach for calling recombination in trios and dyads is summarized in Figure 3.8.



**Figure 3.8 Overview of methods for calling recombination in trios and dyads**

### 3.7.2 Calling Stage of Origin in Trios

Determining the stage of origin of trisomy 21 in a complete parent-child trio first entails comparing the parents' and their child's genotypes at informative markers on chromosome 21. Briefly, a marker is informative if the parent in which the error was derived (the mother, in our application) is heterozygous and the genotypes of both parents together make it possible to infer whether the parent of origin passed two identical versus two different alleles to the child. In the former case, the marker is scored as "reduced to homozygosity" (R), and in the latter "not reduced to homozygosity" (N). Assuming there are no genotyping errors, recombination events are indicated by changes in zygosity between adjacent markers.

After scoring markers (STRs and SNPs) on chromosome 21 in this way, we determined the stage of meiotic error by considering the scores of proximal markers (those nearest the

centromere) on the long arm of the chromosome. Assuming no recombination events occur between the centromere and the most proximal marker, the stage of origin is indicated by the zygosity of that proximal marker: N for an MI error, or R for an MII error.

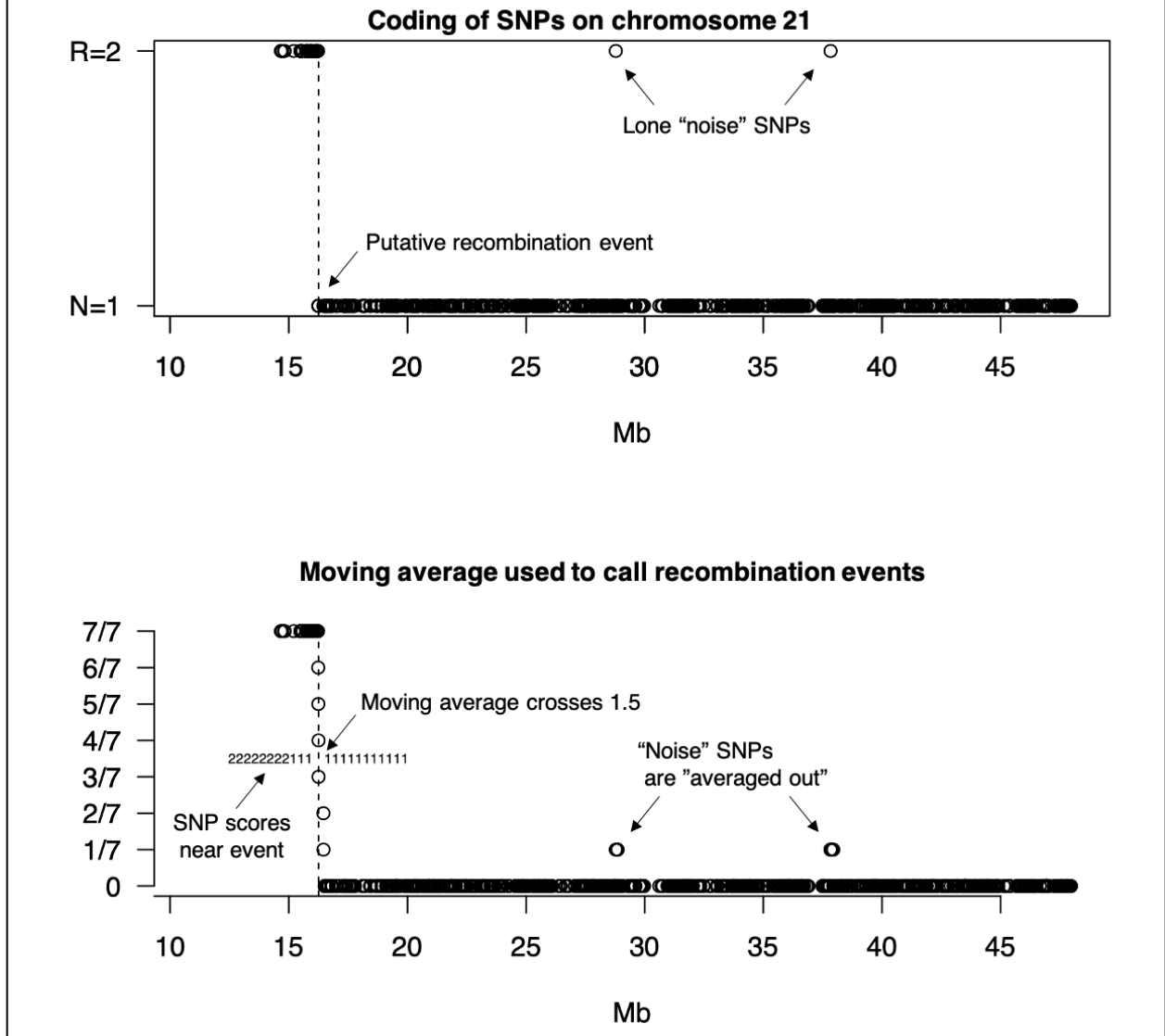
In almost all trios, the most proximal informative marker was a SNP, but in a few cases the stage of origin was called based on an STR. In nine trios, the stage of origin call was determined by a single SNP or STR (that is, the zygosity changed at the very next marker). In the 15 trios where the parent of origin was unclear based on SNP genotypes (due to mosaicism or poor genotyping), no attempt was made to call the stage of origin. These were excluded from our analyses.

### **3.7.3 Calling Recombination Events in Trios**

As described above, recombination events are indicated by changes in zygosity between nearby ordered markers. In the presence of genotyping error, however, not every change in zygosity truly represents recombination. For instance, the pattern RRRRRNRRRRR likely reflects a genotyping error at a single SNP rather than two extremely close recombination events, whereas RRRRRNNNNRRRRR shows stronger evidence of two recombination events (since multiple, independent genotyping errors in a row are less likely). Thus there is a tradeoff between over- and under-calling recombination from changes in SNP scores.

Here we used a simple algorithm to call recombination in the presence of genotyping errors. At each informative SNP in the interior of the chromosome, we considered a moving window centered at that SNP and consisting of the information from seven SNPs: the SNP itself and the three nearest SNPs on either side. SNPs are coded as N=1 and R=2 and then averaged for each window (Figure 3.9).

- Step 1. Code each SNP along chromosome 21q (see top plot)  
 Reduced to homozygosity (R) → 2  
 Not reduced to homozygosity (N) → 1
- Step 2. Calculate moving 7-SNP average at each SNP (see bottom plot)
- Step 3. If the moving average crosses 1.5 between two SNPs, call a recombination event halfway between them



**Figure 3.9 Method for calling recombination events in a full trio**

For the three most informative proximal SNPs on the chromosome arm and the three most telomeric SNPs, we had to use a different method to calculate the seven-SNP-window score. We made the assumption that the chance for a more extreme recombinant (i.e., one closer to the centromere or one beyond the most telomeric SNP) was highly unlikely. Thus, we simply counted

the most proximal (telomeric) SNP more than once toward these averages. This procedure yielded a moving average at each SNP, for the purpose of smoothing out “noise” due to genotyping errors.

Recombination events were called when the moving average of each window crossed a threshold of 1.5 (equivalent to the local 7-SNP “majority vote” changing). The exact location of the recombination was called as halfway between the two informative SNPs flanking the event. In the nine trios where the stage of error was called on the basis of a single SNP or STR, recombination was also called on the basis of that single proximal SNP or STR. In the two cases where an STR was used, the recombination location was approximated with the position of the most proximal SNP.

To call multiple recombinants along the chromosome using a window size of seven SNPs effectively placed a lower bound of four SNPs for the plausible distance between two recombination events. This bears out the intuition (arrived at through trial and error) that patterns such as RRRRRNNNNRRRRR and RRRRRNNNRNRRRRR represent true double-recombination events (the latter with some “noise” due to genotyping error), but that RRRRRNNNRNRRRRR may not.

The number of recombination events for each trio was also recorded. In two trios with excessive “noise” in the scores, recombination events could not be called with confidence. In all, recombination events were called for 613 trios. Plots showing the recombination profiles are included in the Supporting Information.



### 3.7.4 Overview of Inference in Dyads

In addition to the 613 full trios in which we successfully called recombination events, we also attempted to call stage of origin and recombination events in 94 mother-child dyads. Our approach is similar to that used above for trios: scoring informative markers and calculating moving averages. The analysis differs because the absence of paternal genotype information, which results in less informative marker scoring. While markers can still be scored as R in dyads (e.g., mother=AT; child=TTT), the N state cannot be directly inferred by comparing the genotypes of mother and child (e.g., mother=AT; child=AAT). However, this state (X) provides partial information. Calling stage and recombination therefore entails making plausible inference about the true state when X is observed. For instance, when we observe RRRRXRRRR or RRRRXXXXRRRR, we can assume that X masks R in the first case with more confidence than in the second case. Observing RRRRXXXXXXXXXXXXXXXXRRRR, it becomes likely that there is a double recombinant event and that the series of Xs is masking N. Using this intuition, we tend to infer a state of N where the density of X is high relative to R.

Because our study focuses on maternal nondisjunction and recombination in oocytes, we did not attempt to determine stage of origin or call recombination events in father-child dyads.

### 3.7.5 Calling Stage of Origin in Dyads

Our method for calling stage of meiotic error has been described elsewhere [18]. Briefly, following the logic described above, we used the ratio of the number of X to the number of R SNPs near the centromere to predict the true zygosity there (R or N). The number of SNPs to include and the classification threshold for this ratio were treated as parameters to be optimized. The full

trios were used as a training set to obtain the optimal parameters, since masking the father's genotypes in a trio yields a mother-child dyad (with stage of meiotic origin determined above).

### **3.7.6 Calling Recombination Events in Dyads**

To call recombination events in the 94 mother-child dyads, we used an algorithm combining the logic used for calling recombination in the full trios and the logic for calling stage of origin in the dyads (see Figure 3.8).

First, informative SNPs on chromosome 21 were scored as R or X in each dyad. As in the trios, we formed a window centered at each informative SNP. Using the coding  $R=2$  and  $X=3$ , a moving average for a window was calculated using each SNP (as before, the most proximal and distal SNP were counted more than once in order to provide a full window around telomeric/centromeric "edge" SNPs).

As above, we sought to call a recombination event whenever the moving average crossed a threshold between adjacent SNPs (using average position of the two SNPs flanking the event). The number of SNPs in the window and the threshold were treated as parameters to be optimized, the number of possible thresholds depending on the number of SNPs in the window.

We used the full trios as a training set to optimize these parameters, treating the recombination calls already obtained for the trios as the truth. Each trio represents a mother-child dyad plus a father, so by masking the father's genotypes and coding SNPs ( $R=2$ ,  $X=3$ ) we obtained 613 pseudo-dyads with "true" recombination calls.

For each window size (ranging from 7 to 41 SNPs, using odd numbers of SNPs), a moving average was calculated across the chromosome. SNPs near the ends of the chromosome were given

extra weight, as above. For each possible threshold, recombination events were called in each trio (halfway between any two SNPs where the threshold was crossed).

Because an interval that is reduced to homozygosity in a dyad will be represented by a more or less random streak of 2s and 3s, the moving average may fluctuate rapidly even over a short distance. Therefore we used several filtering steps to avoid oversensitivity in calling tight double recombination events and recombination events near the ends of the chromosome. Tight double recombination calls were filtered out if they were separated by fewer than seven informative SNPs or less than 500kb. More stringent filtering was applied to tight double recombination calls representing a middle segment not reduced to homozygosity (i.e. a middle segment with a high proportion of partially informative SNPs coded X). These were filtered out unless supported by a streak of 40 consecutive X SNPs. Recombination calls based on inferring non-homozygosity near the ends of the chromosome were filtered out if not supported by a streak of 12 X SNPs at the centromeric end or 40 X SNPs near the telomeric end. We applied these filtering steps both in training our algorithm in the pseudo-dyads and in making final recombination calls for the dyads.

For each possible pair of parameter values (i.e., the number of SNPs in a window and the threshold for calling a recombinant), the recombination calls were scored numerically on each pseudo-dyad. The “best” pair of parameter values was defined to be the one that maximized the total score across all pseudo-dyads in the training set. We calculated the score as follows.

		Truth (based on trio)	
		k = 0 events	k > 0 events
Calls (based on pseudo-dyad)	m = 0 calls	+1	-k
	m > 0 calls	-m	$k_1 - k_2 - (m - k_1)$
<b>Legend: <math>k = k_1 + k_2</math>, where <math>k_1</math> = number of true events correctly detected and <math>k_2</math> = number of true events not detected</b>			

**Figure 3.10 Scoring recombination calls in the training set trios**

For simplicity, all types of correct/incorrect recombinant calls were weighed equally in a trio (see Figure 3.10). In the case of a pseudo-dyad with  $k=0$  true recombination events, a score of +1 was added if the algorithm correctly called 0 events in that “dyad”; otherwise -1 was deducted for each incorrectly-called event. In the case of  $k > 1$  true events, +1 was awarded for each of the  $k$  events detected by a call within 1 Mb, -1 was deducted for each of the  $k$  events not detected by a call within 1 Mb, and -1 was deducted for each additional event erroneously called. Figure 3.11 shows an example of calling recombination and scoring in a masked dyad. Plots showing recombination-calling in all 613 pseudo-dyads are included in the Supporting Information.

- Step 0. Call recombination events in full trios, as described above; treat these calls as the "truth" (see Figure A2)  
 Use the full trio's genotypes to code each SNP along chromosome 21q (see top plot)
- Step 1. Ignore the father's genotype and code each SNP in the pseudo-dyad on 21q (see second plot)  
 Reduced to homozygosity (R) → 2  
 Partially informative (X) → 3
- Step 2. For different window sizes and possible thresholds:  
 Call recombination events where the threshold is crossed (see third plot, showing optimal threshold)  
 Grade the recombination calls against the "truth" calculated in step 0

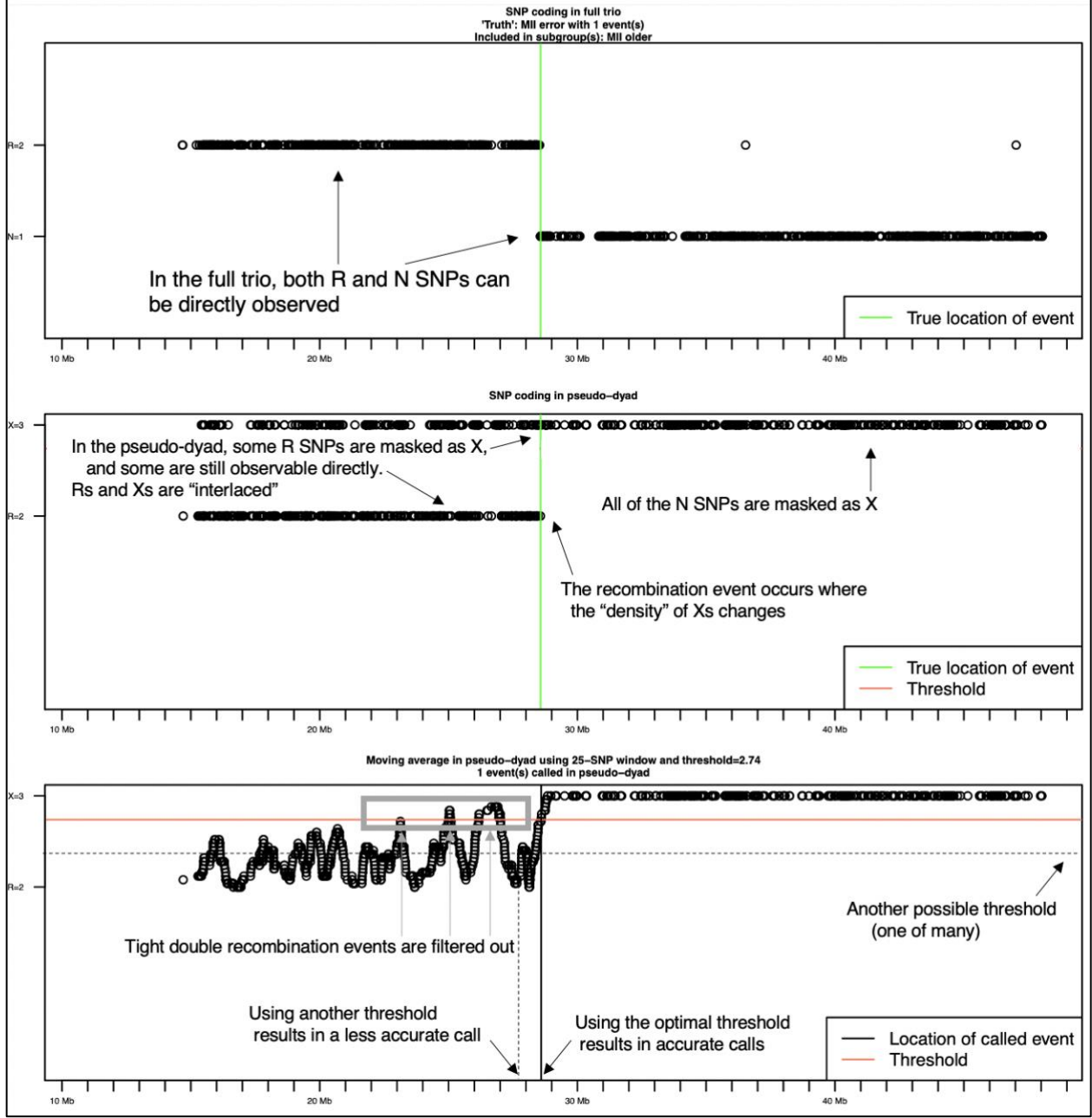


Figure 3.11 Method for calling recombination events in a masked-trio dyad



### **3.8 ACKNOWLEDGMENTS**

The authors thank the Origins of Aneuploidy Research Consortium for informative and fruitful discussions.

## 4.0 CONCLUSION

### 4.1 SUMMARY OF FINDINGS

This dissertation has three aims: performing a GWAS and candidate gene study to find genetic risk factors for maternal meiotic nondisjunction of chromosome 21, calling recombination events on chromosome 21 in trios and dyads, and then dissecting the genetic associations, stratifying by recombination profile (as well as maternal age and meiotic stage of error). These aims were successful. The initial study ([Chapter 2](#)) found plausible associations with candidate genes such as the synaptonemal complex gene *SYCP2* and the cohesin subunit gene *RAD21L* as well as suggestive GWAS associations near genes such as *VEGFA*, *SLC39A10*, *AURKC*, and *MYO10*. The results were consistent with the existence of MI- and MII-specific risk factors as well as shared risk factors. In the second aim ([Chapter 3, Appendix](#)) we called recombination events on chromosome 21 in trios and found that those calls can be used in turn to train an accurate algorithm for calling recombination in mother-child dyads. In the third aim ([Chapter 3](#)) we performed stratified analyses of the candidate and GWAS loci in subgroups of mothers defined by recombination profile, age, and meiotic stage of error. The results are consistent with models in which some variants may, for example, confer general nondisjunction risk across all subgroups, some may act broadly within MI or MII, and some could cause or exacerbate particular “susceptible” recombination patterns. While these analyses are not designed to prove or disprove any models or mechanisms, they could prove to be a first step toward describing the genetic architecture of meiotic nondisjunction.



## 4.2 STRENGTHS AND LIMITATIONS

This study is the first genome-wide association study of meiotic nondisjunction, a surprisingly common occurrence in humans with important implications for health and reproduction. We found evidence that plausible loci may confer risk of meiotic errors, possibly in a mechanism-specific manner. In particular, the candidate gene analysis successfully combined GWAS data with biological insights to begin dissecting the genetics of meiotic nondisjunction.

Given the modest sample size of this study, it is unsurprising that the GWAS did not reach genome-wide significance (whereas some candidate loci did meet a less stringent Bonferroni threshold). As in any genetic association study, the results of our genome-wide, candidate gene, and follow-up stratified analyses need to be replicated in an independent cohort.

Our work here with a population-based cohort complements previous studies of aneuploidy in oocytes, which have often been performed in the context of assisted reproduction and therefore may not be representative of the general population. However, we only analyzed common SNPs in a mostly European-background group. Therefore, in addition to studying rare and structural variants, we should also study meiotic nondisjunction in other populations.

## 4.3 FUTURE ANALYSES

In addition to replication studies of the same phenotype in other populations and studying rarer variation, several other follow-up approaches may be feasible, some more purely epidemiological and others that could combined standard genetic epidemiology with exciting new techniques.

First, we plan further genome-wide association studies of recombination phenotypes in this cohort. In those analyses, genetic variants will be tested for association with “susceptible” recombination patterns on chromosome 21 and the number of recombination events on the chromosome. An interesting extension would be to phase the genotypes in parent-child trios, which would enable recombination-calling on the other chromosomes, in turn allowing us to investigate genome-wide recombination in mothers who experienced meiotic nondisjunction of chromosome 21.

Second, recent molecular and imaging advances have partially overcome some of the technical and ethical challenges of studying the origins of aneuploidy in humans. These include the novel procedure of “meiomapping”, which combines maternal autosomal DNA with SNP array data from the first and second polar bodies along with oocyte (or early embryo) [144]. Analysis of all of the products of meiosis in addition to maternal DNA accomplishes several important tasks. First, it enables inference of the mechanism in an instance of meiosis that has resulted in aneuploidy (see Figure 1.1). Second, it allows inference of meiotic stage of error and genome-wide recombination patterns. Together, these full-meiotic-product data have yielded important new insights that could not be fully addressed through mathematical modeling, such as disentangling rate estimates for specific error mechanisms and establishing selection for recombinant chromatids at MII [111, 112].

The genetic epidemiology approach of this dissertation could be extended to the context of meiomapping by performing more refined stratified genome-wide analyses to identify risk loci for specific error mechanisms (possibly further stratified by recombination pattern and/or maternal age). By stratifying into what may be the most biologically meaningful subsets and by accessing meiotic products relatively early (before a large proportion of aneuploidies are spontaneously

aborted), such studies may have better power to find associations. This may also enable us to identify interactions of genetics with age for mechanisms in which maternal age plays a role.

Recent research has corroborated the underlying the U-shaped aneuploidy vs. age curve, where the mechanism of aneuploidies for MI errors in the younger mothers (the left side of the “U”) are less clear than those in older mothers [113]. If sufficient samples were available, a meiomapping and GWAS approach in this group could help us infer the error mechanisms and genetic risk factors for MI errors in young women. A limitation of the meiomapping approach would be that such data, often obtained from assisted reproduction contexts, may not be representative of human meioses in general. Another complementary approach may be enabled by advances in single-cell RNA-seq that enable sensitive detection of mitotic or meiotic aneuploidies in embryos [145].

Finally, another direction would be to perform molecular genetics studies of loci identified here (particularly GWAS loci). For example, if knockout/knockdown of homologous genes in model organisms causes aberrant recombination or higher aneuploidy rates, this would constitute important corroboration of our results and could help to associate particular genes and variants with specific error mechanisms.

## BIBLIOGRAPHY

- [1]. Hassold T, Hall H, Hunt P. The origin of human aneuploidy: where we have been, where we are going. *Hum Mol Genet.* 2007;16 Spec No. 2:R203-8.
- [2]. Hassold T, Hunt P. To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet.* 2001;2(4):280-91.
- [3]. Nagaoka SI, Hassold TJ, Hunt PA. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nat Rev Genet.* 2012;13(7):493-504.
- [4]. Franasiak JM, Forman EJ, Hong KH, Werner MD, Upham KM, Treff NR, et al. Aneuploidy across individual chromosomes at the embryonic level in trophoctoderm biopsies: changes with patient age and chromosome structure. *J Assist Reprod Genet.* 2014;31(11):1501-9.
- [5]. Hassold T, Chiu D, Yamane JA. Parental origin of autosomal trisomies. *Ann Hum Genet.* 1984;48(2):129-44.
- [6]. Sherman SL, Allen EG, Bean LJH. Maternal Age and Oocyte Aneuploidy: Lessons Learned from Trisomy 21. In: Schlegel PN, Fauser BC, Carrell DT, Racowsky C, editors. *Biennial Review of Infertility: Volume 3.* New York, NY: Springer New York; 2013. p. 69-85.
- [7]. Freeman SB, Allen EG, Oxford-Wright CL, Tinker SW, Druschel C, Hobbs CA, et al. The National Down Syndrome Project: design and implementation. *Public Health Rep.* 2007;122(1):62-72.
- [8]. Lamb NE, Sherman SL, Hassold TJ. Effect of meiotic recombination on the production of aneuploid gametes in humans. *Cytogenet Genome Res.* 2005;111(3-4):250-5.
- [9]. Lamb NE, Freeman SB, Savage-Austin A, Pettay D, Taft L, Hersey J, et al. Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II. *Nat Genet.* 1996;14(4):400-5.
- [10]. Lamb NE, Yu K, Shaffer J, Feingold E, Sherman SL. Association between maternal age and meiotic recombination for trisomy 21. *Am J Hum Genet.* 2005;76(1):91-9.
- [11]. Oliver TR, Feingold E, Yu K, Cheung V, Tinker S, Yadav-Shah M, et al. New insights into human nondisjunction of chromosome 21 in oocytes. *PLoS Genet.* 2008;4(3):e1000033.
- [12]. Oliver TR, Middlebrooks CD, Tinker SW, Allen EG, Bean LJ, Begum F, et al. An examination of the relationship between hotspots and recombination associated with chromosome 21 nondisjunction. *PLoS One.* 2014;9(6):e99560.
- [13]. Oliver TR, Tinker SW, Allen EG, Hollis N, Locke AE, Bean LJ, et al. Altered patterns of multiple recombinant events are associated with nondisjunction of chromosome 21. *Hum Genet.* 2012;131(7):1039-46.
- [14]. Lamb NE, Feingold E, Savage A, Avramopoulos D, Freeman S, Gu Y, et al. Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjunction of chromosome 21. *Hum Mol Genet.* 1997;6(9):1391-9.
- [15]. Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, et al. Common and low-frequency variants associated with genome-wide recombination rate. *Nat Genet.* 2014;46(1):11-6.
- [16]. Antonarakis SE, Petersen MB, McInnis MG, Adelsberger PA, Schinzel AA, Binkert F, et al. The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms. *Am J Hum Genet.* 1992;50(3):544-50.

- [17]. Lin Y, Tseng GC, Cheong SY, Bean LJ, Sherman SL, Feingold E. Smarter clustering methods for SNP genotype calling. *Bioinformatics*. 2008;24(23):2665-71.
- [18]. Chernus JM, Allen EG, Zeng Z, Hoffman ER, Hassold TJ, Feingold E, et al. A candidate gene analysis and GWAS for genes associated with maternal nondisjunction of chromosome 21. *PLoS Genet*. 2019;15(12):e1008414.
- [19]. Chernus JM, Sherman SL, Feingold E. Analyses stratified by maternal age and recombination further characterize genes associated with maternal nondisjunction of chromosome 21. *Prenat Diagn*. 2021;41(5):591-609.
- [20]. Hook EB, Mutton DE, Ide R, Alberman E, Bobrow M. The natural history of Down syndrome conceptuses diagnosed prenatally that are not electively terminated. *Am J Hum Genet*. 1995;57(4):875-81.
- [21]. Bolcun-Filas E, Schimenti JC. Genetics of meiosis and recombination in mice. *Int Rev Cell Mol Biol*. 2012;298:179-227.
- [22]. Cahoon CK, Hawley RS. Regulating the construction and demolition of the synaptonemal complex. *Nat Struct Mol Biol*. 2016;23(5):369-77.
- [23]. Zickler D, Kleckner N. Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harb Perspect Biol*. 2015;7(6).
- [24]. Moses MJ. Chromosomal structures in crayfish spermatocytes. *J Biophys Biochem Cytol*. 1956;2(2):215-8.
- [25]. Fawcett DW. The fine structure of chromosomes in the meiotic prophase of vertebrate spermatocytes. *J Biophys Biochem Cytol*. 1956;2(4):403-6.
- [26]. Wojtasz L, Daniel K, Roig I, Bolcun-Filas E, Xu H, Boonsanay V, et al. Mouse HORMAD1 and HORMAD2, two conserved meiotic chromosomal proteins, are depleted from synapsed chromosome axes with the help of TRIP13 AAA-ATPase. *PLoS Genet*. 2009;5(10):e1000702.
- [27]. Daniel K, Lange J, Hached K, Fu J, Anastassiadis K, Roig I, et al. Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nat Cell Biol*. 2011;13(5):599-610.
- [28]. Cole F, Keeney S, Jasin M. Evolutionary conservation of meiotic DSB proteins: more than just Spo11. *Genes Dev*. 2010;24(12):1201-7.
- [29]. Sansam CL, Pezza RJ. Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. *FEBS J*. 2015;282(13):2444-57.
- [30]. de Vries SS, Baart EB, Dekker M, Siezen A, de Rooij DG, de Boer P, et al. Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev*. 1999;13(5):523-31.
- [31]. Edlmann W, Cohen PE, Kane M, Lau K, Morrow B, Bennett S, et al. Meiotic pachytene arrest in MLH1-deficient mice. *Cell*. 1996;85(7):1125-34.
- [32]. Kneitz B, Cohen PE, Avdievich E, Zhu L, Kane MF, Hou H, Jr., et al. MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes Dev*. 2000;14(9):1085-97.
- [33]. Lipkin SM, Moens PB, Wang V, Lenzi M, Shanmugarajah D, Gilgeous A, et al. Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat Genet*. 2002;31(4):385-90.
- [34]. Plug AW, Peters AH, Keegan KS, Hoekstra MF, de Boer P, Ashley T. Changes in protein composition of meiotic nodules during mammalian meiosis. *J Cell Sci*. 1998;111 ( Pt 4):413-23.
- [35]. Manhart CM, Alani E. Roles for mismatch repair family proteins in promoting meiotic crossing over. *DNA Repair (Amst)*. 2016;38:84-93.

- [36]. Middlebrooks CD, Mukhopadhyay N, Tinker SW, Allen EG, Bean LJ, Begum F, et al. Evidence for dysregulation of genome-wide recombination in oocytes with nondisjoined chromosomes 21. *Human molecular genetics*. 2014;23(2):408-17.
- [37]. Brown AS, Feingold E, Broman KW, Sherman SL. Genome-wide variation in recombination in female meiosis: a risk factor for non-disjunction of chromosome 21. *Hum Mol Genet*. 2000;9(4):515-23.
- [38]. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol*. 2010;34(6):591-602.
- [39]. Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, Zackai EH, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet*. 2010;19(7):1263-75.
- [40]. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res*. 2006;16(9):1136-48.
- [41]. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012;44(6):642-50.
- [42]. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genetics*. 2006;2(12):e190.
- [43]. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
- [44]. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
- [45]. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008;17(R2):R122-8.
- [46]. Shaffer JR, Wang X, Feingold E, Lee M, Begum F, Weeks DE, et al. Genome-wide association scan for childhood caries implicates novel genes. *J Dent Res*. 2011;90(12):1457-62.
- [47]. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal Human Genetic*. 1993;52:506-16.
- [48]. Kerstann KF, Feingold E, Freeman SB, Bean LJ, Pyatt R, Tinker S, et al. Linkage disequilibrium mapping in trisomic populations: Analytical approaches and an application to congenital heart defects in Down syndrome. *Genet Epidemiol*. 2004;27(3):240-51.
- [49]. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*. 2005;95(3):221-7.
- [50]. Lee J, Hirano T. RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. *J Cell Biol*. 2011;192(2):263-76.
- [51]. Ishiguro K, Kim J, Fujiyama-Nakamura S, Kato S, Watanabe Y. A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. *EMBO Rep*. 2011;12(3):267-75.
- [52]. Gutierrez-Caballero C, Herran Y, Sanchez-Martin M, Suja JA, Barbero JL, Llano E, et al. Identification and molecular characterization of the mammalian alpha-kleisin RAD21L. *Cell Cycle*. 2011;10(9):1477-87.

- [53]. Herran Y, Gutierrez-Caballero C, Sanchez-Martin M, Hernandez T, Viera A, Barbero JL, et al. The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. *EMBO J*. 2011;30(15):3091-105.
- [54]. Watanabe Y, Nurse P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature*. 1999;400(6743):461-4.
- [55]. Klein F, Mahr P, Galova M, Buonomo SB, Michaelis C, Nairz K, et al. A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell*. 1999;98(1):91-103.
- [56]. Minase G, Miyamoto T, Miyagawa Y, Iijima M, Ueda H, Saijo Y, et al. Single-nucleotide polymorphisms in the human RAD21L gene may be a genetic risk factor for Japanese patients with azoospermia caused by meiotic arrest and Sertoli cell-only syndrome. *Hum Fertil (Camb)*. 2017;20(3):217-20.
- [57]. Pezza RJ, Voloshin ON, Volodin AA, Boateng KA, Bellani MA, Mazin AV, et al. The dual role of HOP2 in mammalian meiotic homologous recombination. *Nucleic Acids Res*. 2014;42(4):2346-57.
- [58]. Kerzendorfer C, Vignard J, Pedrosa-Harand A, Siwiec T, Akimcheva S, Jolivet S, et al. The Arabidopsis thaliana MND1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *J Cell Sci*. 2006;119(Pt 12):2486-96.
- [59]. Domenichini S, Raynaud C, Ni DA, Henry Y, Bergounioux C. Atmnd1-delta1 is sensitive to gamma-irradiation and defective in meiotic DNA repair. *DNA Repair (Amst)*. 2006;5(4):455-64.
- [60]. Zhao W, Sung P. Significance of ligand interactions involving Hop2-Mnd1 and the RAD51 and DMC1 recombinases in homologous DNA repair and XX ovarian dysgenesis. *Nucleic Acids Res*. 2015;43(8):4055-66.
- [61]. Lee DH, Goodarzi AA, Adelmant GO, Pan Y, Jeggo PA, Marto JA, et al. Phosphoproteomic analysis reveals that PP4 dephosphorylates KAP-1 impacting the DNA damage response. *EMBO J*. 2012;31(10):2403-15.
- [62]. Yoon YS, Lee MW, Ryu D, Kim JH, Ma H, Seo WY, et al. Suppressor of MEK null (SMEK)/protein phosphatase 4 catalytic subunit (PP4C) is a key regulator of hepatic gluconeogenesis. *Proc Natl Acad Sci U S A*. 2010;107(41):17704-9.
- [63]. Kim BR, Seo SH, Park MS, Lee SH, Kwon Y, Rho SB. sMEK1 inhibits endothelial cell proliferation by attenuating VEGFR-2-dependent-Akt/eNOS/HIF-1alpha signaling pathways. *Oncotarget*. 2015;6(31):31830-43.
- [64]. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six Novel Loci Associated with Circulating VEGF Levels Identified by a Meta-analysis of Genome-Wide Association Studies. *PLoS Genet*. 2016;12(2):e1005874.
- [65]. McFee RM, Cupp AS. Vascular contributions to early ovarian development: potential roles of VEGFA isoforms. *Reprod Fertil Dev*. 2013;25(2):333-42.
- [66]. Robinson RS, Woad KJ, Hammond AJ, Laird M, Hunter MG, Mann GE. Angiogenesis and vascular function in the ovary. *Reproduction*. 2009;138(6):869-81.
- [67]. Kim AM, Vogt S, O'Halloran TV, Woodruff TK. Zinc availability regulates exit from meiosis in maturing mammalian oocytes. *Nat Chem Biol*. 2010;6(9):674-81.
- [68]. Kim AM, Bernhardt ML, Kong BY, Ahn RW, Vogt S, Woodruff TK, et al. Zinc sparks are triggered by fertilization and facilitate cell cycle resumption in mammalian eggs. *ACS Chem Biol*. 2011;6(7):716-23.
- [69]. Lisle RS, Anthony K, Randall MA, Diaz FJ. Oocyte-cumulus cell interactions regulate free intracellular zinc in mouse oocytes. *Reproduction*. 2013;145(4):381-90.

- [70]. Dieterich K, Zouari R, Harbuz R, Vialard F, Martinez D, Bellayou H, et al. The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. *Hum Mol Genet.* 2009;18(7):1301-9.
- [71]. Fellmeth JE, Ghanaim EM, Schindler K. Characterization of macrozoospermia-associated AURKC mutations in a mammalian meiotic system. *Hum Mol Genet.* 2016;25(13):2698-711.
- [72]. Adriaenssens T, Mazoyer C, Segers I, Wathlet S, Smits J. Differences in collagen expression in cumulus cells after exposure to highly purified menotropin or recombinant follicle-stimulating hormone in a mouse follicle culture model. *Biol Reprod.* 2009;80(5):1015-25.
- [73]. Baumann C, Wang X, Yang L, Viveiros MM. Error-prone meiotic division and subfertility in mice with oocyte-conditional knockdown of pericentrin. *J Cell Sci.* 2017;130(7):1251-62.
- [74]. Holubcova Z, Blayney M, Elder K, Schuh M. Human oocytes. Error-prone chromosome-mediated spindle assembly favors chromosome segregation defects in human oocytes. *Science.* 2015;348(6239):1143-7.
- [75]. Begum F, Chowdhury R, Cheung VG, Sherman SL, Feingold E. Genome-Wide Association Study of Meiotic Recombination Phenotypes. *G3 (Bethesda).* 2016;6(12):3995-4007.
- [76]. Mao S, Wu F, Cao X, He M, Liu N, Wu H, et al. TDRP deficiency contributes to low sperm motility and is a potential risk factor for male infertility. *Am J Transl Res.* 2016;8(1):177-87.
- [77]. Chen PJ, Huang YS. CPEB2-eEF2 interaction impedes HIF-1alpha RNA translation. *EMBO J.* 2012;31(4):959-71.
- [78]. Kurihara Y, Tokuriki M, Myojin R, Hori T, Kuroiwa A, Matsuda Y, et al. CPEB2, a novel putative translational regulator in mouse haploid germ cells. *Biol Reprod.* 2003;69(1):261-8.
- [79]. Weber KL, Sokac AM, Berg JS, Cheney RE, Bement WM. A microtubule-binding myosin required for nuclear anchoring and spindle assembly. *Nature.* 2004;431(7006):325-9.
- [80]. Samwer M, Dehne HJ, Spira F, Kollmar M, Gerlich DW, Urlaub H, et al. The nuclear F-actin interactome of *Xenopus* oocytes reveals an actin-bundling kinesin that is essential for meiotic cytokinesis. *EMBO J.* 2013;32(13):1886-902.
- [81]. Mogessie B, Schuh M. Actin protects mammalian eggs against chromosome segregation errors. *Science.* 2017;357(6353).
- [82]. Liu QY, Lei JX, Sikorska M, Liu R. A novel brain-enriched E3 ubiquitin ligase RNF182 is up regulated in the brains of Alzheimer's patients and targets ATP6V0C for degradation. *Mol Neurodegener.* 2008;3:4.
- [83]. Nectoux J, Fichou Y, Rosas-Vargas H, Cagnard N, Bahi-Buisson N, Nusbaum P, et al. Cell cloning-based transcriptome analysis in Rett patients: relevance to the pathogenesis of Rett syndrome of new human MeCP2 target genes. *J Cell Mol Med.* 2010;14(7):1962-74.
- [84]. Goel P, Manning JA, Kumar S. NEDD4-2 (NEDD4L): the ubiquitin ligase for multiple membrane proteins. *Gene.* 2015;557(1):1-10.
- [85]. Low LH, Chow YL, Li Y, Goh CP, Putz U, Silke J, et al. Nedd4 family interacting protein 1 (Ndfip1) is required for ubiquitination and nuclear trafficking of BRCA1-associated ATM activator 1 (BRAT1) during the DNA damage response. *J Biol Chem.* 2015;290(11):7141-50.
- [86]. Xu C, Fan CD, Wang X. Regulation of Mdm2 protein stability and the p53 response by NEDD4-1 E3 ligase. *Oncogene.* 2015;34(3):281-9.
- [87]. Hodges CA, Revenkova E, Jessberger R, Hassold TJ, Hunt PA. SMC1beta-deficient female mice provide evidence that cohesins are a missing link in age-related nondisjunction. *Nat Genet.* 2005;37(12):1351-5.



- [88]. Murdoch B, Owen N, Stevense M, Smith H, Nagaoka S, Hassold T, et al. Altered cohesin gene dosage affects Mammalian meiotic chromosome structure and behavior. *PLoS Genet.* 2013;9(2):e1003241.
- [89]. Hwang G, Sun F, O'Brien M, Eppig JJ, Handel MA, Jordan PW. SMC5/6 is required for the formation of segregation-competent bivalent chromosomes during meiosis I in mouse oocytes. *Development.* 2017;144(9):1648-60.
- [90]. Ward A, Hopkins J, McKay M, Murray S, Jordan PW. Genetic Interactions Between the Meiosis-Specific Cohesin Components, STAG3, REC8, and RAD21L. *G3 (Bethesda).* 2016;6(6):1713-24.
- [91]. Biswas U, Hempel K, Llano E, Pendas A, Jessberger R. Distinct Roles of Meiosis-Specific Cohesin Complexes in Mammalian Spermatogenesis. *PLoS Genet.* 2016;12(10):e1006389.
- [92]. Agostinho A, Manneberg O, van Schendel R, Hernandez-Hernandez A, Kouznetsova A, Blom H, et al. High density of REC8 constrains sister chromatid axes and prevents illegitimate synaptonemal complex formation. *EMBO Rep.* 2016;17(6):901-13.
- [93]. McNicoll F, Stevense M, Jessberger R. Cohesin in gametogenesis. *Curr Top Dev Biol.* 2013;102:1-34.
- [94]. Zickler D, Kleckner N. Meiotic chromosomes: integrating structure and function. *Annu Rev Genet.* 1999;33:603-754.
- [95]. Geisinger A, Benavente R. Mutations in Genes Coding for Synaptonemal Complex Proteins and Their Impact on Human Fertility. *Cytogenet Genome Res.* 2016;150(2):77-85.
- [96]. Broman K, Murray J, Sheffield V, White R, Weber J. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American Journal Human Genetic.* 1998;63:861-9.
- [97]. Cheung VG, Burdick JT, Hirschmann D, Morley M. Polymorphic variation in human meiotic recombination. *American Journal Human Genetic.* 2007;80(3):526-30.
- [98]. Fledel-Alon A, Leffler EM, Guan Y, Stephens M, Coop G, Przeworski M. Variation in human recombination rates and its genetic determinants. *PLoS One.* 2011;6(6):e20321.
- [99]. Hou Y, Fan W, Yan L, Li R, Lian Y, Huang J, et al. Genome analyses of single human oocytes. *Cell.* 2013;155(7):1492-506.
- [100]. Brieno-Enriquez MA, Cohen PE. Double trouble in human aneuploidy. *Nat Genet.* 2015;47(7):696-8.
- [101]. MacLennan M, Crichton JH, Playfoot CJ, Adams IR. Oocyte development, meiosis and aneuploidy. *Semin Cell Dev Biol.* 2015;45:68-76.
- [102]. Wassmann K. Sister chromatid segregation in meiosis II: deprotection through phosphorylation. *Cell Cycle.* 2013;12(9):1352-9.
- [103]. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. *Science.* 2008.
- [104]. Chowdhury R, Bois PR, Feingold E, Sherman SL, Cheung VG. Genetic analysis of variation in human meiotic recombination. *PLoS Genet.* 2009;5(9):e1000648.
- [105]. Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, et al. Sequence Variants in the RNF212 Gene Associate with Genomewide Recombination Rate. *Science.* 2008.
- [106]. Qiao H, Prasada Rao HB, Yang Y, Fong JH, Cloutier JM, Deacon DC, et al. Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nat Genet.* 2014;46(2):194-9.

- [107]. Oliver TR, Middlebrooks C, Harden A, Scott N, Johnson B, Jones J, et al. Variation in the Zinc Finger of PRDM9 is Associated with the Absence of Recombination along Nondisjoined Chromosomes 21 of Maternal Origin. *J Down Syndr Chromosom Abnorm*. 2016;2(2).
- [108]. McCoy RC, Demko Z, Ryan A, Banjevic M, Hill M, Sigurjonsson S, et al. Common variants spanning PLK4 are associated with mitotic-origin aneuploidy in human embryos. *Science*. 2015;348(6231):235-8.
- [109]. Capalbo A, Hoffmann ER, Cimadomo D, Ubaldi FM, Rienzi L. Human female meiosis revised: new insights into the mechanisms of chromosome segregation and aneuploidies from advanced genomics and time-lapse imaging. *Hum Reprod Update*. 2017;23(6):706-22.
- [110]. Capalbo A, Bono S, Spizzichino L, Biricik A, Baldi M, Colamaria S, et al. Sequential comprehensive chromosome analysis on polar bodies, blastomeres and trophoblast: insights into female meiotic errors and chromosomal segregation in the preimplantation window of embryo development. *Hum Reprod*. 2013;28(2):509-18.
- [111]. Ottolini CS, Newnham L, Capalbo A, Natesan SA, Joshi HA, Cimadomo D, et al. Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat Genet*. 2015;47(7):727-35.
- [112]. Tyc KM, McCoy RC, Schindler K, Xing J. Mathematical modeling of human oocyte aneuploidy. *Proc Natl Acad Sci U S A*. 2020;117(19):10455-64.
- [113]. Gruhn JR, Zielinska AP, Shukla V, Blanshard R, Capalbo A, Cimadomo D, et al. Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science*. 2019;365(6460):1466-9.
- [114]. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria 2017.
- [115]. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-75.
- [116]. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7.
- [117]. Nishimura T, Sugiura K, Naito K. A-kinase anchor protein 1 (AKAP1) regulates cAMP-dependent protein kinase (PKA) localization and is involved in meiotic maturation of porcine oocytes. *Biol Reprod*. 2013;88(4):85.
- [118]. Zhang RN, Pang B, Xu SR, Wan PC, Guo SC, Ji HZ, et al. The CXCL12-CXCR4 signaling promotes oocyte maturation by regulating cumulus expansion in sheep. *Theriogenology*. 2018;107:85-94.
- [119]. Chen Z, Kang X, Wang L, Dong H, Wang C, Xiong Z, et al. Rictor/mTORC2 pathway in oocytes regulates folliculogenesis, and its inactivation causes premature ovarian failure. *J Biol Chem*. 2015;290(10):6387-96.
- [120]. Wang SB, Liu YP, Zhang ZH, Wang ZK, Xu YX, Wang ZC. Temporal and spatial expression profiles of Frizzled 3 in the ovary during the estrous cycle. *Genet Mol Res*. 2016;15(1).
- [121]. Segers I, Adriaenssens T, Smits J. Expression patterns of poliovirus receptor, erythrocyte protein band 4.1-like 3, regulator of g-protein signaling 11, and oxytocin receptor in mouse ovarian cells during follicle growth and early luteinization in vitro and in vivo. *Biol Reprod*. 2012;86(1):1-11.

- [122]. Cao Z, Gao D, Xu T, Zhang L, Tong X, Zhang D, et al. Circular RNA profiling in the oocyte and cumulus cells reveals that circARMC4 is essential for porcine oocyte maturation. *Aging* (Albany NY). 2019;11(18):8015-34.
- [123]. Kong BY, Duncan FE, Que EL, Kim AM, O'Halloran TV, Woodruff TK. Maternally-derived zinc transporters ZIP6 and ZIP10 drive the mammalian oocyte-to-egg transition. *Mol Hum Reprod*. 2014;20(11):1077-89.
- [124]. Nguyen AL, Schindler K. Specialize and Divide (Twice): Functions of Three Aurora Kinase Homologs in Mammalian Oocyte Meiotic Maturation. *Trends Genet*. 2017;33(5):349-63.
- [125]. Ojima F, Saito Y, Tsuchiya Y, Ogoshi M, Fukamachi H, Inagaki K, et al. Runx3 regulates folliculogenesis and steroidogenesis in granulosa cells of immature mice. *Cell Tissue Res*. 2019;375(3):743-54.
- [126]. Awe JP, Byrne JA. Identifying candidate oocyte reprogramming factors using cross-species global transcriptional analysis. *Cell Reprogram*. 2013;15(2):126-33.
- [127]. de Castro RO, Previato L, Goitea V, Felberg A, Guiraldelli MF, Filiberti A, et al. The chromatin-remodeling subunit Baf200 promotes homology-directed DNA repair and regulates distinct chromatin-remodeling complexes. *J Biol Chem*. 2017;292(20):8459-71.
- [128]. Shi X, Sun X, Liu M, Li D, Aneja R, Zhou J. CEP70 protein interacts with gamma-tubulin to localize at the centrosome and is critical for mitotic spindle assembly. *J Biol Chem*. 2011;286(38):33401-8.
- [129]. Sakthianandeswaren A, Parsons MJ, Mouradov D, MacKinnon RN, Catimel B, Liu S, et al. MACROD2 Haploinsufficiency Impairs Catalytic Activity of PARP1 and Promotes Chromosome Instability and Growth of Intestinal Tumors. *Cancer Discov*. 2018;8(8):988-1005.
- [130]. Woolner S, O'Brien LL, Wiese C, Bement WM. Myosin-10 and actin filaments are essential for mitotic spindle function. *J Cell Biol*. 2008;182(1):77-88.
- [131]. Sandquist JC, Larson ME, Hine KJ. Myosin-10 independently influences mitotic spindle structure and mitotic progression. *Cytoskeleton* (Hoboken). 2016;73(7):351-64.
- [132]. Kwon M, Bagonis M, Danuser G, Pellman D. Direct Microtubule-Binding by Myosin-10 Orients Centrosomes toward Retraction Fibers and Subcortical Actin Clouds. *Dev Cell*. 2015;34(3):323-37.
- [133]. Brieno-Enriquez MA, Moak SL, Holloway JK, Cohen PE. NIMA-related kinase 1 (NEK1) regulates meiosis I spindle assembly by altering the balance between alpha-Adducin and Myosin X. *PLoS One*. 2017;12(10):e0185780.
- [134]. Hou S, Li N, Zhang Q, Li H, Wei X, Hao T, et al. XAB2 functions in mitotic cell cycle progression via transcriptional regulation of CENPE. *Cell Death Dis*. 2016;7(10):e2409.
- [135]. Onyango DO, Howard SM, Neherin K, Yanez DA, Stark JM. Tetratricopeptide repeat factor XAB2 mediates the end resection step of homologous recombination. *Nucleic Acids Res*. 2016;44(12):5702-16.
- [136]. Budna J, Chachula A, Kazmierczak D, Rybska M, Ciesiolka S, Bryja A, et al. Morphogenesis-related gene-expression profile in porcine oocytes before and after in vitro maturation. *Zygote*. 2017;25(3):331-40.
- [137]. Kulus M, Kranc W, Jeseta M, Sujka-Kordowska P, Konwerska A, Ciesiółka S, et al. Cortical Granule Distribution and Expression Pattern of Genes Regulating Cellular Component Size, Morphogenesis, and Potential to Differentiation are Related to Oocyte Developmental Competence and Maturation Capacity In Vivo and In Vitro. *Genes* (Basel). 2020;11(7).
- [138]. Prochazkova B, Komrskova P, Kubelka M. CPEB2 Is Necessary for Proper Porcine Meiotic Maturation and Embryonic Development. *Int J Mol Sci*. 2018;19(10).

- [139]. Einarson MB, Cukierman E, Compton DA, Golemis EA. Human enhancer of invasion-cluster, a coiled-coil protein required for passage through mitosis. *Mol Cell Biol*. 2004;24(9):3957-71.
- [140]. Wang SB, Xing BS, Yi L, Wang W, Xu YX. Expression of Frizzled 2 in the mouse ovary during oestrous cycle. *J Anim Physiol Anim Nutr (Berl)*. 2010;94(4):437-45.
- [141]. Bothun AM, Woods DC. Dynamics of WNT signaling components in the human ovary from development to adulthood. *Histochem Cell Biol*. 2019;151(2):115-23.
- [142]. Murakami H, Lam I, Huang PC, Song J, van Overbeek M, Keeney S. Multilayered mechanisms ensure that short chromosomes recombine in meiosis. *Nature*. 2020;582(7810):124-8.
- [143]. Qiao H, Rao H, Yun Y, Sandhu S, Fong JH, Sapre M, et al. Impeding DNA Break Repair Enables Oocyte Quality Control. *Mol Cell*. 2018;72(2):211-21 e3.
- [144]. Ottolini CS, Capalbo A, Newnham L, Cimadomo D, Natesan SA, Hoffmann ER, et al. Generation of meiomaps of genome-wide recombination and chromosome segregation in human oocytes. *Nat Protoc*. 2016;11(7):1229-43.
- [145]. Starostik MR, Sosina OA, McCoy RC. Single-cell analysis of human embryos reveals diverse patterns of aneuploidy and mosaicism. *Genome Res*. 2020;30(6):814-25.