

**Transcriptomic Analysis Reveals a Molecular Synergy Between
the Maternally Deposited Chromatin Modifying Enzymes: SPR-5, MET-2
and the Zygotically Derived Nucleosome Remodeling Deacetylase
Complex Component MEP-1**



Jovan S. Brockett

05/04/2020

Oglethorpe University

Committee Members: Dr. Karen Schmeichel, Dr. Onur Birol, Dr. Lea Alford

Introduction:

Epigenetics is the study of phenotypic changes that do not involve alterations of genomic DNA nucleotide sequences. These phenotypic alterations result from differential gene expression that are governed by mechanisms “above” (from the Greek “epi”) the genetic code. Differential gene expression is essential for the development of various cell lineages within an embryo and into adulthood. Developmental signals induce differential gene expression not only by initiating transcription via transcription factors and cis-regulatory elements, but also by causing changes in chromatin structure to provide steric access to the core DNA sequences. (D’Urso et al., 2014) Fluctuations in gene expression defines cell identity and their potential for differentiation. Like mutations and allele variants, epigenetic changes can also be maintained through subsequent cell divisions, even in the absence of the initial developmental signals.

One of the best studied examples of the interplay between gene expression and epigenetic changes is the developing of *Drosophila melanogaster* embryo, where expression of homeotic genes is established by transient expression of the segmentation transcription factors (D’Urso et al 2014). The access of these transcription factors to DNA sequences is a result of altering chromatin structure via epigenetic mechanisms (Stephens et al., 2018). After these factors release their bind of DNA, the expression patterns of many genes, including the homeotic genes, is maintained through many cell divisions (D’Urso et al., 2014). Gene regulation and maintenance is also achieved through post-translational epigenetic mechanisms such as RNA silencing/activation via siRNAs, sRNAs, and microRNAs. Increasingly it appears that among the most influential epigenetic changes regulating gene expression are covalent

modifications to either the DNA itself (DNA methylation/acetylation) or post-translational covalent modifications of histones H2A, H2B, H3 and H4. These changes result in an alteration of the timing and amplitude of when genes are expressed, that ultimately dictate the ability of canonical transcriptional machinery to read the DNA code. In this way, modification of chromatin structure by chromatin modifying enzymes are critical determinants of cell fate in a developing embryo.

Chromatin has two opposing states: euchromatin where DNA is transcriptionally active (or accessible by transcriptional machinery), and heterochromatin where DNA is transcriptionally inactive (cannot be accessed by transcriptional machinery) (Bannister et al., 2011). The extent to which the chromatin transitions between these two states is a crucial determinant of gene expression. The basic repeating structural (and functional) unit of chromatin is the nucleosome, which contains eight histone proteins and about 146 base pairs of DNA (Van Holde, 1988; Wolffe, 1999, Annunziato et al., 2008). Histones can be enzymatically modified on their chains of amino acid tails by the addition or removal of acetyl, methyl, or phosphate groups (Fischle et al., 2005). The most typical modifications of gene expression are a result of post-translational histone modifications (PTMs) on amino acid chains (tails) of the histones (Bowman et al 2015). Likewise, these marks are reversible and can be added and removed enzymatically in response to developmental signals (Greer et al 2012). Together this system is intrinsically plastic, as each mark can be altered in response to environmental and cellular stimuli.

Histone tail modifications contribute to the control of gene expression by influencing chromatin compaction and accessibility of cis regulatory elements such as

gene promoters to transcriptional machinery (Bannister et al 2011). The histone marks provide a 3-dimensional chemical signature that recruits enzymes which can further activate or repress gene expression (Dong et al 2013). These enzymes are known as readers, writers, and erasers. Together they work in concert with one another to properly regulate a gene (Zhang et al., 2015). The quality and quantity of histone modifications and the degree to which those marks activate or repress chromatin constitutes the “histone code,” a code that is far from being understood.

This histone code is massively complex and the understanding of which is the reason this is such an enormous project comparable to that of the human genome project (Riveria et al., 2013). For example, the tails of the four standard histones: H2A, H2B, H3, and H4 can be modified at different sites with different modifications; some marks act together while other marks are mutually exclusive (Zhang et al 2015). For example, histone 3 tails contain nineteen lysines which are known to be either mono, di, or tri methylated (Karachenstev et al., 2006). In theory, there are approximately 44 million histones in the human genome and if these modifications are independent of each other, this allows a potential for 280 billion different lysine methylation patterns in every cell. Lysine acetylation, methylation and arginine methylation, increase global accessibility of DNA which open up heterochromatin.

Similarly, this does not even include modifications of other histones such as H2A, H2B, or H4.

In theory, every nucleosome in a cell could potentially have a different set of modifications. Such a possibility is mindboggling and raises the question of whether evolution has selected for common reproducible patterns of histone modifications. A

study of 40 histone modifications across human gene promoters found over 4000 different combinations of histone modifications used (Chih et al., 2005). Despite the enormity of possibilities, patterns were discovered including a set of 17 histone modifications that are present together at over 3000 genes (Wang et al 2008). This indicates possible collaboration between multiple chromatin modifying proteins to perform a common task of either repression or activation of genes. Therefore, recognizable and repeated patterns of histone modifications do occur, but they are very intricate. We currently have detailed biochemical understanding of the importance of a relatively small number of histone modifications. Thus, elucidation of the readers, writers, and erasers in establishing and utilizing the histone code requires additional research to come.



Mechanisms of Histone Methylation and Acetylation

Histone methylation and acetylation, two writer “marks”, occur on all basic amino acids, arginines, lysines, and histidines (Smith et al., 2010). This process involves methyl or acetyl groups being transferred or removed to amino acid residues of histone protein tails in nucleosomes (Rechsteiner et al 2010). Regulation of histone methylation and acetylation is achieved through a family of so called “writer” enzymes: those that catalyze addition of methyl groups (histone methyl transferases), those that catalyze the addition of acetyl groups (histone acetyl transferases), so called erasers: those that catalyze removal of methyl groups (histone demethylases), and those that catalyze the removal of acetyl groups (histone deacetylases) (Greer et al 2012). The targets of the enzymes are R-groups of basic amino acids that are found in histone tails (Bannister et al 2011). The most extensively studied histone methylation and acetylation sites on the

histone tails include those with lysine on histone 3, lysines 4, 9, 27, and 36 (i.e. H3K4, H3K9, H3K27, H3K36)

In the case of deacetylation, an acetyl group is removed from the histone tail giving the histone a more positive charge. This increases the affinity of the N-terminus of histones to the negative charge phosphate groups of DNA, making the DNA tightly wound upon histone octamers rendering it inaccessible from transcriptional machinery (Erlar et al., 2014). This is the process by which euchromatin, or transcriptionally accessible/active DNA, is physically altered to heterochromatin, or transcriptionally inaccessible/repressed DNA.

While acetylation is generally considered an active mark, the effects of histone methylation is not as predictable. Histone methylation can trigger chromatin activation or repression depending what amino acid residues are methylated. For example, the methylation to H3K4me2 results in active gene transcription, however the methylation to H3K9me2 results in the repression of gene transcription (Miller et al, 2013). The difference lies in the ability of each of these chemically modified surfaces to recruit enzymes to the methyl marks. Methyltransferases, acetyltransferases, demethylases and deacetylases all work in harmony with each other to regulate gene expression by placing active and repressive marks at the right time and the right place throughout the genome (Kerr et al., 2014). Collectively, data from many biological systems indicate epigenetic regulation of gene expression is essential to ensure appropriate cellular development and it appears to govern cell fate decisions immediately after sperm and egg fuse to form a zygote all the way through adulthood.

C. *elegans* as a model to study epigenetic reprogramming

The mechanistic complexities of epigenetic reprogramming through histone modifying enzymes has begun to be systematically explored using tractable genetic model organisms such as *Caenorhabditis elegans*. *C. elegans* is well suited as an experimental model due to its short life cycle, fast reproductive capabilities, transparent structure, predictable life cycle and known genome sequence.

Studies in *C. elegans* make it easy to apply principles seen in the orthologues of *C. elegans* to the mouse model so we can further investigate these findings in a vertebrate system. Due to extensive research done upon *C. elegans*, there is a plethora of information and genetic tools available to facilitate studies of molecular genetics. The *C. elegans* genome has been fully sequenced for decades and is supported by publicly accessible databases (i.e. Wormbase, Wormenricher) that exemplifies wild-type phenotypes at particular larval stages, full genome sequencing, gene locations on a chromosome, libraries of null allele mutants, and libraries of gene constructs.

C. elegans can be engineered with an array of phenotypically marked balancer chromosomes that can be used to determine chromosomal recombination and ensure predicted sorting of desirable mutant alleles. Balancer chromosomes allow for potentially lethal mutations to be maintained throughout generations when in a heterozygous state with the balancer chromosomes. This is especially imperative when trying to maintain mutations that affect maternally deposited gene products. *C. elegans* respond effectively to RNA interference (RNAi) using fast and simple methods to knockout any gene of interest in a single generation (Fire et al., 1998). In addition, *C. elegans* do not undergo DNA methylation, eliminating a confounding epigenetic factor from investigation. This is especially important as it allows for a more

acute focus on methylation of histones, helping to uncover the significance of this dynamic epigenetic mark on regulating global gene expression during development.

Epigenetic reprogramming by SPR-5 and MET-2

In order for *C. elegans* to properly develop from an embryo into an organism, a series of intricate and perfectly timed steps must occur. *C. elegans* have an invariant lineage, thus fate mapping when lineage tracing cell lines is easier compared to other model organisms. The invariant lineage of this model organism creates an easier outlook on defining germline vs. somatic cell distinction in that after initial cellular division of a zygote results in two primordial germ cells. These cells are destined to become the germline and do not revert back into a somatic cell type. Likewise, at this stage, every other cell is predetermined to be of somatic lineage and is not supposed to become a germ cell. This is important when trying to identify the cellular defects that result from impaired epigenetic machineries (Wu et al., 2012).

Some of the first molecular determinants driving development happen before fertilization when maternally deposited transcripts and proteins are deposited into the oocyte (Katz et al., 2009). Two of these maternally deposited factors are histone modifying enzymes, SPR-5 and MET-2. SPR-5 is a H3K4me2 demethylase and MET-2 is a H3K9me2 methyltransferase (Katz et al., 2009). Both enzymes collaborate to repress expression of oocyte and sperm specific genes in the zygote after fertilization (Katz et al., 2009). They also facilitate the repression of all genes that were left active from the mother during this time (Kerr et al., 2014).

Germline specific genes are initially turned on via H3K4 methylation at the promoter regions of the oocyte and sperm specific genes by activating complexes such as COMPASS (Shilatifard et al., 2012). After fertilization, SPR-5 and MET-2 work together to revert cell fate from a terminal differentiated cell to a pluripotent state (“blank slate”). Once reprogrammed in this way, cells in the embryo can then proliferate and differentiate into any specific cell type (Kerr et al 2014). In mutants lacking either of these enzymes, the transgenerational inheritance of H3K4me2 or lack of H3K9me2 results in progressive sterility in progeny over many generations (approx. 30) (Katz et al., 2009). This progressive sterility is a result of accumulation of active marks at germline genes in the soma that over time are not erased by SPR-5 or marked repressively by MET-2. In double mutants lacking both of these enzymes, the progeny exhibit sterility in a single generation and also a developmental delay phenotype meaning that they are much slower to develop when compared to wild type (Katz et al., 2009). This delay is likely due to the ectopic expression of germline genes in somatic cells causing molecular confusion which prevents somatic cells from dividing properly. *C. elegans* with mutations in *spr-5* and *met-2* corrupt germline-soma distinction. Since both types of programs are simultaneously trying to be turned on, the cellular differentiation process becomes lagged. Evolutionarily speaking, germline formation is the most critical cellular determinant in order to ensure the passing of organisms’ genes to the next generation. Since this is such an important element of cellular differentiation, uncovering how this mechanism is safeguarded by all these various writers, readers, and erasers is critical in

understanding the complexities of organismal development. Thus, understanding how the mechanisms associated with SPR-5 and MET-2 epigenetic reprogramming function, is key to understanding the most basic developmental events.

The role for SPR-5 and MET-2 in regulating early developmental epigenetic events is well established, but other studies have begun to determine the identity of other genes involved in this pathway. For example, recent work has shown that the *mes-4* gene (maternal effect sterile) (along with others, *ash-1*, *set-2*, *jdm-1*, *Jumanji*,) is a focal determinant of early gene regulating decisions (Gaydos et al 2012). MES-4 is now known to be responsible for the methylation of H3K36me₃, which is an active mark (Patel et al., 2012). MES-4 is essential in cell differentiation in distinguishing between germline and somatic cell lines (Gaydos et al 2012). Failure to deposit maternal MES-4 results in sterility in the organism. Distinguishing between germline and soma is achieved through the effect of H3K36 which is specifically found in germline gene loci at transcription start sites (TSS) in germ cells (Patel et al., 2012). MES-4 methylates H3K36me₃ to bookmark germline genes so that they are readily available to be expressed in the germline as soon as the germline begins to form at the 4th larval stage. In the (*spr-5*; *met-2*) double mutants, progeny are born with a plethora of open chromatin since epigenetic reprogramming cannot happen due to the lack of SPR-5 and MET-2 (Carpenter et al., 2020). This epigenetic reprogramming does not occur because SPR-5 is not present to remove H3K4me_{1/2} active marks and MET-2 cannot add H3K9me_{1/2} repressive marks (Carpenter et al., 2020). Since the chromatin in these double mutants is easily accessible MES-4 is free to methylate H3K36me₃ in all cell types because MES-4 does not require transcriptional activation. Essentially MES-4

primes germline genes to be expressed throughout the entirety of the soma where they would usually not be found. This causes much confusion among cellular differentiation and ultimately causes the developmental delay phenotype seen in the *spr-5; met-2* double mutants (referred to going forward as “DM”) (Carpenter et al., 2020). When reducing *mes-4* transcripts via RNAi in the background of the *spr-5; met-2* double mutants, there is a return to wild-type development meaning that the developmental delay phenotype was rescued (Carpenter et al., 2020). The hypothesis supported by these results is that with the removal of MES-4 from the system, germline gene expression ceased in somatic cell precursors allowing for normal somatic cell development. However, this rescue comes at the cost of germline fertility since H3K36me3 activation marks cannot be added in germ cell precursors (Carpenter et al., 2020). After the maternally deposited chromatin modifying enzymes establish an epigenetic ground state (Kerr et al., 2014), the need for the zygotic genome to continue maintenance of the initial establishment is required in order to ensure proper development (Schulz et al., 2019).

Unhavaithaya et al (2002) discovered a member of a nucleosome remodeling complex called MEP-1. MEP-1 is a zygotic gene product that is expressed sometime soon after fertilization and was suggested to be one of the controllers of gene expression of germline genes by these germline genes from being activated, however the exact order of action and timing of these events is still not clear (Unhavaithaya et al., 2002). What we do know is MEP-1 is a component of a nucleosome remodeling deacetylase (NuRD) like complex called MEC. NuRD complexes are repressive in their nature, consistent with a role in policing inappropriate

germline gene expression in the soma. These NuRD complexes are highly conserved between species which further illustrates their importance in proper development of any organism. They play a role in the synMuvB developmental pathway playing an essential role in the controlling of oncogenesis genes in humans. Mutations in these synMuvB pathway gene regulators can result in cancer and tumors (Fu et al., 2011). Likewise, these complexes share similar roles between species; that is to further reinforce germline vs. somatic cellular distinction. MEC is comprised of LET-418, HDA-1, and MEP-1 all of which work together to deacetylate histone amino acid residues (repressive modification) in order to prevent germline gene expression in somatic cells. RNAi knockdown of the *mep-1* results in a larval arrest (unable to develop past the L1 stage) in wild-type *C. elegans* (Unhavaithaya et al., 2002). A knockdown of *mes-4* in the *mep-1* knockdown mutant resulted in a rescue in the larval 1 arrest similar to that of a rescue in *spr-5; met-2 (mes-4)* triple mutants (Carpenter et al., 2020). Similarly, the *spr-5; met-2* double mutants resembled an arrest to that of the *mep-1* single mutants.

Previous research used RNAi to reduce *mep-1* transcript levels in the background of the DM (*spr-5; met-2*). The prediction was that circumstantial evidence would support the potential collaboration between maternal and zygotic systems to reinforce proper development. The results of this experiment resulted in an L1 arrest phenotypically worse than *spr-5; met-2* alone or *mep-1* knockdown alone (Chavez 2019). This exacerbated larval arrest implicates a synergistic effect of these chromatin modifiers and further molecular confusion causing problems during normal developmental events (Chavez 2019). With this data we suspected that the intensified

response is a result of SPR-5, MET-2, and MEP-1 acting upon common gene targets and are therefore reinforcing the same developmental pathway.

Here this synergy is examined further by performing a transcriptomic analysis of L1 *C. elegans* deficient for *mep-1*, *spr-5; met-2* (DM), and *spr-5; met-2; mep-1* (DM(*mep-1*)). The rationale for this experiment is, if these proteins are converging on similar gene targets, there must be overlap in genes misexpressed in N2(*mep-1*), *spr-5; met-2* and *spr-5; met-2; mep-1* mutants. Similarly, we would expect that the expression of those same sets of misexpressed genes should be further exacerbated in an additive fashion as the number of mutant genes increase. Lastly, you would expect of those misexpressed genes, some should be germline genes. Alternatively, if they have differing gene targets then perhaps there is a unique function of MEP-1 which has yet to be discovered. Either outcome contributes to our understanding of the complex interplay between these epigenetic players as they dictate early developmental cell fate decisions.

GLETHORPE
UNIVERSITY

MATERIALS AND METHODS:

Strains. All *Caenorhabditis elegans* strains were grown and maintained at 20° C under standard conditions, as previously described (Brenner, 1974). The *C. elegans spr-5 (by101)(I)* strain was provided by (Lakowski et al 2003). The N2 Bristol wild-type (WT) and FX30208 *tmC27 [unc-75(tmls1239)] (I)* strain was obtained by the Caenorhabditis Genetics Center (CGC). The *tmls1239* chromosome balancer completely covers the *spr-5* gene on chromosome 1. The *tmC27* allele contains a wild type version of the *spr-5* gene that is linked to a green fluorescent pharynx protein gene (GFP). The GFP emits a 488-nm illumination in the pharynx of the animal if the gene is present. The *qC1 [qls26 (lag2::GFP + rol-6(su1006))](III)* strain was obtained from the W. Kelly (Bowman et al., 2013) and crossed to *met-2 (n4256)(III)* to maintain *met-2(n4256)(III)* as heterozygotes. The *qc1* “roller” allele on the chromosome 3 balancer manifests phenotypically at larval stage 3 or 4, thus genotyping must be confirmed in earlier larval stages via *lag2* GFP found in the Z1 and Z4 primordial somatic cells. The *spr-5 (by101)(I)/tmC27[unc-75(tmls1239)](I); met-2 (n4256) (III)/qC1 [qls26 (lag2::gfp+ rol 6(su1006))](III)* strain was re-created for this study to maintain *spr-5 (by101)(I); met-2 (n4256)(III)* double-mutant animals as balanced heterozygotes. (Carpenter et al., 2020). Going forward, these mutants will be referred to as Double heterozygotes (DH). Animals homozygous for *spr-5* and *met-2* mutations will be referred to as double mutants (DM). The selection scheme for DMs from DHs is illustrated in Figure 1.

Strain Maintenance. *C. elegans* strains were maintained on OP-50 bacteria spotted on NGM plates. NGM agar plates seeded with OP-50 bacteria were allowed to dry at least overnight. All strains were moved onto new seeded NGM plates approximately every 4

days to avoid plate starvation, which can stress animals and cause changes to their epigenetic status (Phillips et al., 2019). The N2 Bristol wild-type (WT) strain was maintained by moving 3 L4 worms onto a new plate. The DH worms were maintained by transferring 4 GFP pharynx positive L4s onto seeded NGM plates. (Animals homozygous for *tmc27* (*tmc27[unc-75(tmls1239)]*) (I) balancer also had green pharynxes, however, they were two-times brighter than the heterozygous animals). The parent generation used to produce the worms used for the RNAi experiments: *spr-5/spr-5;met-2/qc1*, are homozygous for *spr-5* mutation and heterozygous for *met-2*. This strain was not maintained more than 1 generation because of the transgenerational effects seen with *spr-5/spr-5* homozygous animals by Katz et al., 2009, Kerr et al., 2014 and Greer et al., 2014. Without the *tmC27* balancer allele paired with *spr-5*, the effects seen in these mutants could carry over multiple generations resulting in transgenerational sterility seen in those previous studies.

The *qc1* balancer [*qls26 (lag2::gfp+ rol 6(su1006))*](III) codes for a rolling animal that does not move in the sinusoidal pattern. The *qc1* allele is dominant and having two copies of the *qc1* allele on chromosome 3 is lethal. Therefore, all rolling animals are heterozygous at the 3rd allele with wild type *met-2*. Due to the possibility of allele recombination during reproduction, creating multiple plates of the homozygous animals rather than the heterozygous animals resulted in a higher chance of producing offspring homozygous for both *spr-5* and *met-2* mutations. Homozygous *spr-5; met-2* animals, were non-roller (moving in normal sinusoidal pattern) and non-green pharynx since they do not have the *tmC27* or *qc1* alleles (Figure 1). These animals without *spr-5* and *met-2*

were used in the RNAi experiments and were scored as the “F1” generation. They experience zygotic defects seen in F1 and their offspring (F2).

RNAi Knockdown. RNAi by feeding was carried out using clones from the Ahringer library as described previously by Kamath et al. (2003). Feeding experiments were performed on RNAi plates (NGM plates containing 100 ug/ml ampicillin, 0.4mM IPTG, and 12.5ug/ml tetracycline). The bacterial strain used for RNAi is HT115(DE3), a modified form of *E. coli*. DE3 refers to the IPTG-inducible T7 RNA polymerase cassette, (including a tetracycline resistance gene), that is engineered into the HT115 genomic DNA (Conte et al., 2017). The RNAi expression vector, pL4440, was transformed into HT115 (DE3) contains two T7 promoters in opposite orientations and an ampicillin resistance selection gene. The target gene sequences are cloned into the restriction enzyme site (BglII) between the two T7 promoters. Bacterial cells that do not grow on the NGM plates (i.e. those that have lost the essential T7 gene) are consequentially eradicated. Three different RNAi vectors were used: pL4440, pL4440-*pos-1*, and pL4440-*mep-1* (Figure 2) (Kamath et al., 2003).

In order to deliver the RNAi to the worm cultures, 5mL (LB+ 100µL Amp) was inoculated with a fresh batch from -80°C stock. Liquid cultures were grown at 37°C for 16 hours; cultures were stored (4°C) and used for up to 7 days. Bacterial RNAi cultures were seeded by spotting plain RNAi plates (NGM plates containing 100 ug/ml ampicillin, 0.4mM IPTG, and 12.5ug/ml tetracycline) (3-spots per plate). Seeded RNAi plates were allowed 48 hours to dry and induce dsRNA transcription via IPTG stimulation.

Developmental Delay Assay. 4 DH parents were used to create an F0 generation of the double mutants. F0 DH Animals were allowed to produce F1 offspring of DM (*spr-5/spr-5; met-2/met-2*) which were used for the RNAi experiments. Genotype was confirmed via fluorescence-microscopy by lack of green-pharynx (tmC27) and lack of roller phenotype (*qc1*). The lack of these fluorescent markers indicates an animal homozygous for the mutant alleles: *spr-5* and *met-2*. The F1s were subjected to each RNAi treatment and after 48 hours were transferred to a new RNAi plate of the same bacterial type. The now gravid F1 parents underwent a synchronized lay where each parent could lay their embryos with an allotted amount of time (typically 4-6 hours). Scoring of total progeny and developmental delay was conducted via light-microscopy following removal of F1 parents. A total of 48 hours was allowed for the now F2 generation offspring to develop and image comparing relative size to wild type.

RNAi experiment. When starting with DHs, there is only a 1/8 chance of obtaining the correct genotype used for the experiments. Therefore, amplification of *spr-5* (*by101*)(I)/tmC27[unc-75(tmIs1239)](I); *met-2* (*n4256*) (III)/*qc1* [*qls26* (*lag2::gfp+ rol6*(*su1006*))](III) strain plates were generated to create large scale cultures of *spr-5*(*by101*)/*spr-5*(*by101*) (I)/*met-2*(*n4256*)/*qc1* (III) used for the experiments. From here, 30 F1 L4s (*spr-5/spr-5;met-2/met-2*) worms were placed on RNAi plates (*pL4440*, *pL4440-pos-1*, *pL4440-mep-1*) for a total of 2 plates percondition.

After 48 hrs the then gravid adults were moved to plain RNAi plates where they were allotted time to lay their embryos. F1 worms were then removed from plates and sacrificed and F2 embryos were allowed to develop. After 8 hours the embryos began to hatch; all F2 progeny arrested at the L1 stage in the absence of food. F2 L1 larvae were

then washed off the RNAi plate using 700 μ L of M9 buffer solution and pipetted (glass) into a dimpled glass for further washing and counting (Figure 2). From here the L1 isolates were pipetted up into 1.5mL non-stick Eppendorf tubes. Isolates were spun down for 2.5 min. After spin cycle, 600 μ L M9 buffer was aspirated off the top of the tubes (or until 100 μ L M9 is left). Total number of isolates were counted and then resuspended in a final volume of 100 μ L M9. 10 μ L of M9 solution was counted for total number of animals. 250-1000 isolates were recovered per RNAi treatment..

RNA Extraction. All the RNA manipulations were performed in an RNase free environment (Rnase zap, autoclaved tubes, Rnase zapped gloves, etc). Total RNA was isolated using TRIzol reagent (Invitrogen) from 3000 starved L1 larvae born at room temperature (21°C - 22°C) overnight in un-seeded NGM agar plate. Tubes containing approx. 500 isolates in each had 1mL TRIzol/100 μ L of larval suspension, were vortexed 2x for 30 seconds, and were immediately refrozen into liquid nitrogen. 10 μ L of 20mg/mL glycerin solution was added to the solution followed by the addition of 200 μ L bromochloropropane. 0.75 volume isopropanol was added to the aqueous layer and RNA was precipitated overnight at -20°C. Isopropanol was removed and pellet washed with 500 μ L 75% ethanol. RNA was spun down at for 10 min. Final dried RNA pellets were solubilized in 20 μ L dH₂O.

RNA Sequencing. Total RNA was submitted to The Georgia Genomics and Bioinformatics Core (GGBC) <https://dna.uga.edu/>. GGBC operates multiple platforms for short-, long-, and single-molecule sequencing reads (i.e., Illumina MiSeq and NextSeq, PacBio Sequel, and Oxford Nanopore Minlon). GGBC reverse transcribed the isolated

RNA into DNA and PCR amplified the cDNA. GGBC used Next-gen sequencing (NGS) on the cDNA and then sent the results back in .tabular files.

Transcriptomic Analysis. Three biological replicates of each genotype/RNAi combination were analyzed. Every downstream analysis takes into account significance value based off a α -value $p < .05$. Sequencing reads were checked for quality using FastQC (Wingett and Andrews, 2018), filtered using Trimmomatic (Bolger et al., 2014), and remapped to the *C. elegans* transcriptome (ce10, WS220) using HISAT2 (D. Kim et al., 2015). Read count by gene was obtained by FeatureCounts (Liao et al., 2014). Differentially expressed transcripts (significance threshold p-value < 0.05) were determined using DESEQ2 (v.2.11.40.2) (Love et al., 2014). Transcripts per million (TPM) values were calculated from raw data obtained from FeatureCounts output. Subsequent downstream analysis was performed using R with normalized counts and p-values from DESEQ2 (v.2.11.40.2). Data was scaled and hierarchical clustering was performed using the complete linkage algorithm. In the linkage algorithm, distance was measured by calculating pairwise distance. Additionally, Gene Ontology (GO) Pathway analysis was performed using the online platform WormEnrichr (Chen et al., 2013; 469 Kuleshov et al., 2016). An additional heatmap comparison of differentially expressed genes between *spr-5*, *met-2*, and *spr-5; met-2* progeny compared to N2 progeny was generated in Microsoft Excel using Log2 fold change values from the DESEQ2 analysis. Because transcript isoforms were ignored, the data is discussed in terms of “genes expressed” rather than “transcripts expressed.”

FastQC. FastQC provides a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a

modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis. The main functions of FastQC are: import of data from BAM, SAM, or FastQ files (any variant), provide a quick overview to tell you in which areas there may be problems, summarize graphs and tables to quickly assess your data, export results to an HTML based permanent report, and allow offline operation for automated generation of reports without running the interactive application (you do not need to have internet).

Trimmomatic. Trimmomatic is another QC check by removing sequencing adapters that could interfere with and confound downstream analyses. This would result in mapping the wrong gene to the loci or even not find any like complementary sequences.

HISAT2. HISAT2 is aligns sequences to a known genome reference. (ce10)

Feature Counts. In many applications, the key information required for downstream analysis is the number of reads mapping to each genomic feature, for example to each exon or each gene. The process of counting reads is called read summarization. Running feature counts shows the number of matches the sequence has to a given gene on a known genome. The more matches there are the higher the count. Thus, the higher the count is for a particular gene, the more transcripts there are of that specific gene and ultimately the more that gene is being expressed.

DESeq2. DESeq2 is a method for differential analysis of count data gathered from Feature counts. DESeq2 uses shrinkage estimation for dispersions and fold changes to improve stability and interpretability of estimates. Essentially, this program is what

differentiates the amount of matches of gene from normal levels of expression vs. over or under expressed genes.



O G L E T H O R P E
U N I V E R S I T Y

Results:

Loss of *spr-5*; *met-2*; and *mep-1* causes developmental arrest.

Previous work demonstrated that *spr-5*; *met-2* double mutants (DM) exhibit a developmental delay phenotype due to the misexpression of germline genes (Kerr et al. 2014). Similarly, *mep-1* single mutants exhibit a developmental arrest due to the misexpression of germline genes (Unhavaithaya et al., 2002). To examine whether or not the developmental delay in *spr-5*; *met-2* mutants is synergistic with *mep-1* mutants, we synchronized laid N2, *spr-5*, *met-2*, *mep-1* RNAi, *spr-5*; *mep-1*, *met-2*; *mep-1*(RNAi), *spr-5*; *met-2*, and *spr-5*; *met-2*; *mep-1*(RNAi) mutant hermaphrodites and monitored their development from when they hatched to when they were adults. Figure 3 provides representative results seen by Carpenter et al., 2020 and Chavez 2019. 72 hours post synchronized lay, 100% of N2 progeny, 95% of *spr-5* progeny, and 85% of *met-2* progeny were fertile adults (Figure 3a-c, g) (Carpenter et al, 2020). However, *spr-5*; *met-2* progeny experienced a severe developmental delay. 80% of those progenies resembled L2 larvae after 72 hours, the remaining 20% after 72 hours made it to adulthood, but were sterile (Figure 3d, g) (Carpenter et al 2020). Similarly, *spr-5*; *met-2*; *mep-1* progeny exhibited a severe developmental arrest in 100% of progeny scored (Figure 3e, g) (Chavez, 2019). Some of those developmentally arrested progeny exhibited more intense developmental defects that were not present in the DMs such as cavitation in the cuticle near the head (Figure 3f) (Chavez, 2019). The results reported by Carpenter and Chavez were identically reproduced in this study. The *spr-5*; *met-2* mutants were developmentally delayed and the *spr-5*; *met-2*; *mep-1* mutants were

developmentally arrested (data not shown). The fidelity of these experiments indicates the RNAi is potent enough to be a proxy for a genomic mutation, the experiments were technically reproducible, and it was an efficient way to gather biological replicates respect to both quality and quantity (Data not shown). However, the number of progeny from *spr-5; met-2; mep-1* adults began to dwindle when put into high stress environments such as M9, due to their fragile genetic constitution. Thus, slight adjustments to the protocol were made to increase progeny, and thus to increase yield of RNA. Specifically, *spr-5; met-2; mep-1* progeny laid approximately 10 larvae in M9 buffer solution. This modification resulted in the same mutants laying 500+ progeny on empty RNAi plates instead of M9 (Figure 4). Likewise, switching from M9 buffer to an empty (un-seeded) plate carried the same benefits as the M9 buffer such as: arresting isolates at the L1 stage (due to lack of food), and no cross contamination with OP-50 bacteria. This way, the environment and thus potentially the epigenetic status of these organisms were alike in that they were similarly starved in either treatment as soon as they hatched. Using this “starvation” method, RNA-seq analysis would not be confounded due to transcripts produced by the food OP-50 bacteria. L1 larval arrest was a critical feature of this protocol because these larvae do not express any germline genes. Thus, by arresting all larvae at the L1 stage due to lack of food, we compared the transcriptome of the isolates from each treatment to see whether or not germline gene transcription was occurring in soma. If it is, then the misexpression of these genes is due to the mis regulation of germline genes in either of the mutant backgrounds.

Using this improved ‘starvation’ method, we generated 2000 L1 isolates of the

DM (*mep-1*), DM, N2(*mep-1*), and N2(wild-type). Total RNA was extracted from each of these respective isolates and sent off for sequencing. With this data set, comparative transcriptomics were performed by comparing gene sets through analyzing the amount of up/down regulation (log2fold) of differentially regulated genes. RNA-Seq analyses was used opposed to other analyses such as microarray or Coimmunoprecipitation because: RNA-seq incorporates a broader dynamic range enabling more sensitive and accurate measurement of gene expression, it is not limited by prior knowledge - captures both known and novel changes, it can be applied to any species even if reference sequencing is not available, it is a better value often delivering advantages at a comparable or lower price per sample than many arrays, and it is by far the most cited NGS assay. Similarly, it is a powerful bioinformatics tool that can be used to check levels of any transcript throughout the entire genome in an unbiased fashion. (Roach et al 2019)

***spr-5*, *met-2*, *mep-1* deficient organisms share a significant amount of differentially expressed genes.**

It was previously shown that SPR-5 and MET-2 work synergistically to target and repress ectopic expression of germline genes in somatic tissue. (Kerr et al., 2014) Similarly, it was shown that MEP-1 targets germline genes in somatic tissue to achieve the same goal as SPR-5 and MET-2: repressing ectopic expression of germline genes in somatic cells. (Unhavaithaya et al., 2002) To test the extent to which these chromatin modifying proteins may be collaborating to repress common germline gene targets in somatic cells, we performed RNA seq on N2 (wild-type), N2 with *mep-1* RNA interference (RNAi) [N2(*mep-1*)], *spr-5; met-2*, and *spr-5; met-2; mep-1* L1 progeny.

RNA-seq analyses was performed on L1 larvae because this stage is just before the developmental arrest seen in the *spr-5; met-2* mutants. Likewise, any transcription of germline genes at this larval stage indicates misexpression since these animals do not fully develop germlines until the L4 stage (Kimble et al 2005). Similarly, the data was compared to qRT-PCR results seen in Katz et al., 2009 as a baseline to ensure the changes in experimental measures resulted in similar gene expression profiles.

The first phase of transcriptomic analysis was conducted to ensure that the RNA was in high concentration, but also that the RNA's integrity was not compromised due to the presence of RNases during the extraction process and reads were proportional to input of RNA. The example in Figure 5 shows total RNA concentration for the *spr-5; met-2; mep-1* sample were similar to other samples taken. Table 1 shows the RNA integrity number (RIN) indicating the RNA was not degraded, was in high enough concentration, and high-quality enough to proceed with sequencing. Likewise, the indication of the 18S and 28S ribosomal RNA peaks shows that the total RNA is fully intact (Figure 5). 50 paired-end reads (PE) were requested with no less than 25-million reads resulting in a total of 400 million reads for our 12 samples. Paired-end reads are important to ensure that fragment(s) sequenced were sequenced from both ends and not just one end which could cause issues in further downstream analyses. 1x NextSeq High Output flow cell recorded these reads into a downloadable file from the Illumina website. Each paired end read created a new data file, thus RNA-seq.tabular data files were concatenated using command line unix access terminal and were uploaded to Galaxy to undergo a series of quality control analyses. Concatenation was essential to merge all the sequence sets together into one data file.

Command line was used instead of galaxy to concatenate files because command line allows the control of the order of files to be concatenated. After concatenation, FastQ files were analyzed in Galaxy's FASTQC function on each of the 24 files (two for each sample; forward and reverse). The FastQ files were then used in the trimmomatic function to remove any sequence adapters that could skew downstream analyses. Likewise, Trimmomatic removes the disproportionate base sequence content since the adapters are what was causing the problem in FASTQC. After Trimmomatic quality control, <91% of all reads were left (Table 2). Similarly, HISAT2 was used to align the filtered tags from Trimmomatic onto the known *C. elegans* genome database, ce10. All samples had <97.5% alignment indicating that the RNA not only maintained its integrity, but that it was high quality enough to be completely mapped back to a database (Table 2).

The Featurecounts function was then used to map individual reads onto specific genes or exons. This function known as read summation, is what indicates the total amount of transcripts recorded for a particular gene (Table 2). These counts were summarized by using the DESeq2 function which correlates genes together by how much they are being amplified above wild type expression levels (Love et al., 2014). For example, genes that are more upregulated than other genes and close to the same amount would be grouped together. This creates dispersion plots which can be used to assess the number of genes being misregulated and by how much (Figure 6). The figure shows the trend of misregulated genes and that they shift together and towards the fitted line. Most of the genes are being affected in the same way, as indicted by the tight fit of genes near the projected fitted line. It is safe to assume that the RNA results

are not only from the same organism, but that the mutations in these organisms result in a global change of expression in the same direction (either up or down). What we cannot tell from this analysis though is the directionality of changes of expression.

Having established similar gene misexpression profiles, a detailed analysis of patterning changes was explored. How many genes out of the 20,407 in the *C. elegans* genome were misregulated in the *spr-5; met-2; mep-1* mutants vs. *spr-5; met-2, mep-1* knockdown mutants and wildtype. We identified 77 differentially expressed genes (DEGs, $P_{\text{adjusted}} < .05$) in N2(*mep-1*) mutants ns. wildtype, 4,222 DEGs in the *spr-5; met-2* mutants vs. wildtype, and 7,488 DEGs in the *spr-5; met-2; mep-1* mutants vs. wildtype (Table 3). All 72 DEGs that were misexpressed in *mep-1* RNAi vs. wildtype are also misexpressed in common among all different mutant candidates (Figure 7). However, there was a significant ($P_{\text{adjusted}} < .05$) overlap of 3,621 DEGs between *spr-5; met-2* and *spr-5; met-2; mep-1* mutants.

SPR-5, MET-2, and MEP-1 regulate the expression of the same genes.

While the initial analysis showed common gene targets between SPR-5, MET-2 and MEP-1, it was not clear how much these chromatin remodeling proteins targets impact gene expression on these common gene targets. In order to address this, RNA-seq data was analyzed to construct a scatter correlation plot comparing the DEGs in N2(*mep-1*) knockdown, *spr-5; met-2* and *spr-5; met-2, mep-1* mutants (Figure 8). If the effects are additive, then when comparing N2(*mep-1*), *spr-5; met-2* vs. *spr-5; met-2; mep-1* the results should demonstrate a shift of all genes off the 1:1 correlation line towards a higher log₂FC value on the x-axis. The data demonstrate a significant trend

comparing genes upregulated in the N2(*mep-1*) single mutant, where these same genes are even more upregulated in *spr-5; met-2* double mutants, in some cases being amplified 2-5- fold for each gene (Figure 8a). In N2(*mep-1*) vs. *spr-5; met-2; mep-1* mutants, there is a similar trend only now these same genes plus ≈3000 other genes are being even more expressed in the triple mutant than in the *spr-5; met-2* double. (Figure 8b). The data exemplifies not only both of these, but also a significant difference in expression between the DEGs in *spr-5; met-2* double mutants and the DEGs in *spr-5; met-2; mep-1* “triple” mutant (Figure 9). The bar graph demonstrates the striking difference in gene amplification between the two mutant backgrounds.

SPR-5, MET-2, and MEP-1 regulate the repression of germline genes in somatic tissue.

Collectively, the RNA-seq data suggests that SPR-5, MET-2, and MEP-1 work on a significant amount of the same gene targets. If these two pathways are both working to regulate germline gene expression, then the additive effects should be focused on targets implicated in germline formation (ex, P granules, germ plasm, synaptonemal complex). Using WormEnricher, a gene ontology analyses of the DEGs for N2(*mep-1*), *spr-5; met-2* and *spr-5; met-2; mep-1* mutants was conducted (Figure 10).

WormEnricher output is recorded as combined score. Combined score is calculated by combining, significance of misexpressed genes related with a certain cellular function. Thus, the higher the combined score, the more that category of genes is collectively misregulated.

In wildtype, there are transcripts associated with housekeeping genes and somatic cell function (e.g. endoplasmic reticulum formation to make new membranes). Similarly, genes associated with NuRD complexes are also being transcribed in L1 larvae at the time of RNA extraction (Figure 10a). In N2(*mep-1*) mutants, germline associated genes such as P granule and germ plasm began to appear as misexpressed and was among the top candidates of genes being expressed during RNA extraction (Figure 10b). Likewise, in *spr-5; met-2* mutants, transcripts associated with meiosis and germline formation were among the top gene candidates being misregulated in the absence of SPR-5 and MET-2. These candidates consisted of the same genes being misexpressed in N2(*mep-1*) mutants such as genes associated with the formation of P granules, however their combined score is much higher indicating amplified transcripts at gene loci compared to the single mutant (Figure 10c). In *spr-5; met-2; mep-1* mutants, the same kinds of genes were top candidates, however they were even further misregulated than even the *spr-5; met-2* mutants indicating a synergistic effect between the chromatin modifying proteins. Similarly, new previously undetected candidates of germline formation began to arise as top genes being misexpressed, such as genes associated with the formation of the germ plasm (Figure 11d). The shift in ontological characterization of misexpression towards germ-line machineries, supports the trend expected. These candidates appear when their combined score is higher compared to the combined score of other types of genes.

Discussion:

Previous genetic studies suggested an interaction between maternally deposited chromatin modifying enzymes and MEC via synergistic phenotypic amplification of the developmental delay. To explore this genetic interaction further, an RNA-seq transcriptomic analysis was performed on N2 (wildtype) vs. N2(*mep-1*), *spr-5; met-2*, and *spr-5; met-2; mep-1* mutants. To determine if these two machineries, one maternally deposited and one zygotically activated, police the expression of a common set of genes, *mep-1* was knocked down using RNAi into the background of *spr-5; met-2* mutants. In doing so, it was shown that the inhibition of *mep-1* transcripts via RNAi in these *spr-5; met-2* mutants, resulted in a developmental arrest that the organisms never recover from. This implies that organisms that lack maintenance phase (e.g. NuRD complexes/MEC) chromatin modifying proteins in combination with lack of establishment phase chromatin modifying proteins (e.g. SPR-5 and MET-2), results in a phenotype far more severe than either of the mutations alone. This was achieved through devising a modified protocol that allowed for large scale production of L1 isolates and the extraction of RNA of sufficient quality and quantity to enable high-throughput RNA sequencing. Comparative transcriptomics revealed that maternal reprogramming enzymes, SPR-5 and MET-2 work in collaboration with maintenance phase chromatin modifying proteins such as MEC. These collaborators are in fact, under wild-type circumstances, working in an additive fashion to repress the same gene targets.

Descriptive Ontogeny analysis shows germline genes are being increasingly misexpressed in more mutant genotypes. This is consistent with our hypothesis that MEP-1 and members of MEC are working together with maternal reprogramming

enzymes SPR-5 and MET-2, in order to further facilitate the repression of germline genes in somatic tissues.

The exact timing of when MEC and thus MEP-1 facilitates its reinforcing properties during the maintenance phase is not known. However, this finding is consistent with current models of germ line vs. soma distinction demonstrated in Strome et al 2012. Strome et al demonstrates that other germline regulating chromatin modifiers such as MES-4 are key in bookmarking between germline and somatic cells (Figure 11a). MES-4, in the absence of SPR-5 and MET-2, is responsible for the ectopic expression of germline genes in somatic tissue (Strome et al., 2012). Ultimately this causes the developmental delay seen in *spr-5; met-2* double mutants (Carpenter et al., 2020). This results from H3K36me3 at germline gene loci in somatic cell. Since SPR-5 and MET-2 are not present in DMs to shut down germline genes that are left on from the parents, MES-4 is able to access chromatin in a much wider set of targets than compared to wild type. This causes ectopic expression of genes that are not supposed to be transcribed at that time or in that cell type (Figure 11b). MEC may also have an influence on MES-4 H3K36me3 localization since it reinforces the job of SPR-5 and MET-2 during the maintenance phase after transcription from the zygotic genome.

With the elimination of *mep-1* in the background of the *spr-5; met-2* mutants, we believe that MES-4 has even greater access to even more sets of genes and for longer periods of time since now repressive chromatin complexes are eliminated from both the establishment and maintenance phases (Figure 11c). This potentially could lead to more H3K36 methylation at germline gene loci, leading to higher expression levels of germline genes in somatic tissue. Hence, the exacerbated delay and arrest. This

interpretation is supported by both correlation plots. Thus, we believe this simplified model in the context of these 4 chromatin modifying proteins elucidates the intricacies of proper gene regulation to ensure healthy development. Similarly, preliminary data shows other maintenance phase repressive proteins such as the DREAM complex could also be involved in the same synMuvB pathway as MEC, SPR-5, MES-4, and MET-2. (Ogblethorpe Dev. Bio 2020)

Further biochemical assays such as ChIP-Seq and ATAC-Seq are certainly promising candidates to solidify our findings in this study. A ChIP-Seq analysis of MEP-1 in wildtype vs the *spr-5; met-2* double mutants would give insight into the physical similarities between SPR-5, MET-2 and MEP-1. This would correlate the physical location of MEP-1 with the misexpressed genes from our bioinformatic analyses. The expectation is MEP-1 would localize at germline gene loci that are top candidates in the gene ontogeny reports. Similarly, a ChIP-Seq of H3K36me3 in wild-type vs. *N2(mep-1); spr-5; met-2*; and *spr-5; met-2; mep-1* would allow us to answer the question of MES-4's role in the absence of MEP-1. The expectation of this experiment would be a global increase in H3K36me3 at all genes throughout the genome. Also, a ChIP-seq of global acetylation would be interesting to see how much MEP-1 plays a role in the overall acetylation profile between wild type, *N2(mep-1)*, *spr-5; met-2*, and *spr-5; met-2; mep-1*. Given the results of the RNA-seq data, you would expect MEP-1 to play a big role in global genome acetylation, however there could be other NuRD like complexes that assimilate during later stages of development. Lastly, an ATAC-seq assay on the *spr-5; met-2* double mutants vs. *spr-5; met-2; mep-1* triple mutants would shed light into the global changes of the epigenetic landscape within all of these

mutants. This would indicate patterns of MEP-1 that might affect overall accessibility of the genome. All of these potential assays could be correlated with one another and overlaid to affirm the roles of these chromatin modifying proteins with each other and lead to the identification of other machineries involved in regulating germline/soma distinction.

Collectively these findings shed light on disorders in humans known to be associated with defects in epigenetic reprogramming mechanisms. LSD1, (the human orthologue of SPR-5) deficient patients suffer from severe developmental and neurological disorders such as mental retardation. Similarly, individuals deficient in NuRD complexes result in inflammatory diseases caused by the over expression of inflammatory genes such as rheumatoid arthritis and allergies. (Kuo et al 2014) Mutations in NuRD complexes result in the overexpression of oncogenes associated with tumor progression (Glozak et al 2007). Lastly, NuRD complexes play an essential role in laying down memories and products of behavior such as addiction. (Robison 2011) The need to fully understand the function of all of these chromatin modifying proteins involvement in the developmental process both in adults and embryos as well as gene regulation is becoming essential in continuing the progress of modern medicine.

Citations:

- Annunziato, A. (2008) DNA Packaging: Nucleosomes and Chromatin. *Nature Education* 1(1):26
- Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research*. 2011;21(3):381–395.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
- Bowman EA, Bowman CR, Ahn JH, Kelly WG. Phosphorylation of RNA polymerase II is independent of P-TEFb in the *C. elegans* germline. *Development (Cambridge)*. 2013;140(17):3703–3713.
- Bowman GD, Poirier MG. Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews*. 2015;115(6):2274–2295. doi:10.1021/cr500350x
- Carpenter BS, Lee TW, Plott CF, Brockett JS, Katz D. Failure to maternally reprogram histone methylation causes developmental delay due to germline transcription in somatic tissues. *bioRxiv Developmental Biology*. 2020 Jan 23 [accessed 2020 May 3]:2020.01.22.914580.
- Chavez S. *mep-1* interacts with *spr-5* and *met-2* during embryogenesis in *C. elegans*. Oglethorpe Honors Thesis. 2019.
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR, Ma'ayan A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14.
- Chih LL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biology*. 2005;3(10).
- Conte D, MacNei LT, Walhout AJM, Mello CC. RNA Interference in *Caenorhabditis elegans*. *Current Protocols in Molecular Biology*. 2015;2015:26.3.1-26.3.30.
- Dong X, Weng Z. The correlation between histone modifications and gene expression. *Epigenomics*. 2013;5(2):113–116.
- D'Urso A, Brickner JH. Mechanisms of epigenetic memory. *Trends in Genetics*. 2014;30(6):230–236.

Erler J, Zhang R, Petridis L, Cheng X, Smith JC, Langowski J. The Role of Histone Tails in the Nucleosome: A Computational Study. *Biophysical Journal*. 2014;107(12):2911–2922.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *caenorhabditis elegans*. *Nature*. 1998;391(6669):806–811.

Fischle W, Boo ST, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*. 2005;438(7071):1116–1122.

Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S. Antagonism between MES-4 and Polycomb Repressive Complex 2 Promotes Appropriate Gene Expression in *C. elegans* Germ Cells. *Cell Reports*. 2012;2(5):1169–1177.

Glozak MA, Seto E. Histone deacetylases and cancer. *Oncogene*. 2007;26(37):5420–5432.

Greer EL, Shi Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*. 2012;13(5):343–357.

Fu J, Qin L, He T, Qin J, Hong J, Wong J, Liao L, Xu J. The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Research*. 2011;21(2):275–289.

Kamath RS, Ahringer J. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods*. 2003;30(4):313–321.

Karachentsev D, Druzhinina M, Steward R. Free and chromatin-associated mono-, di-, and trimethylation of histone H4-lysine 20 during development and cell cycle progression. *Developmental Biology*. 2007;304(1):46–52.

Katz DJ, Edwards TM, Reinke V, Kelly WG. A *C. elegans* LSD1 Demethylase Contributes to Germline Immortality by Reprogramming Epigenetic Memory. *Cell*. 2009;137(2):308–320.

Kerr SC, Ruppertsburg CC, Francis JW, Katz DJ. SPR-5 and MET-2 function cooperatively to reestablish an epigenetic ground state during passage through the germ line. *Proceedings of the National Academy of Sciences of the United States of America*. 2014 [accessed 2020 Apr 15];111(26):9509–14.

Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*. 2015;12(4):357–360.

Kimble J, Crittenden S. Germline proliferation and its control. *Worm book*. 2005.

Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*. 2016;44(W1):W90–W97.

Kuo C-H, Hsieh C-C, Lee M-S, Chang K-T, Kuo H-F, Hung C-H. Epigenetic regulation in allergic diseases and related studies. *Asia Pacific Allergy*. 2014;4(1):14. doi:10.5415/apallergy.2014.4.1.14

Lakowski B, Eimer S, Göbel C, Böttcher A, Wagler B, Baumeister R. Two suppressors of sel-12 encode C2H2 zinc-finger proteins that regulate presenilin transcription in *Caenorhabditis elegans*. *Development*. 2003;130(10):2117–2128.

Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–930.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014 [accessed 2020 Apr 30];15(12):550.

Miller JL, Grant PA. The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcellular Biochemistry*. 2013; 61:289–317.

Patel T, Tursun B, Rahe DP, Hobert O. Removal of Polycomb Repressive Complex 2 Makes *C. elegans* Germ Cells Susceptible to Direct Conversion into Specific Somatic Cell Types. *Cell Reports*. 2012;2(5):1178–1186.

Phillips NLH, Roth TL. Animal models and their contribution to our understanding of the relationship between environments, epigenetic modifications, and behavior. *Genes*. 2019;10(1).

Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, Wang W, Kimura H, Lieb JD, Strome S. The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS genetics*. 2010 [accessed 2019 Jun 12];6(9):e1001091.

Rivera CM, Ren B. Mapping human epigenomes. *Cell*. 2013;155(1):39–55.

Robison AJ, Nestler EJ. Transcriptional and epigenetic mechanisms of addiction. *Nature Reviews Neuroscience*. 2011;12(11):623–637.

Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*. 2019;20(4):221–234.

Shilatifard A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annual Review of Biochemistry*. 2012;81(1):65–95.

Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2009;1789(1):45–57.

Stephens KE, Miaskowski CA, Levine JD, Pullinger CR, Aouizerat BE. Epigenetic Regulation and Measurement of Epigenetic Changes. *Biological Research for Nursing*. 2013;15(4):373–381.

Unhavaithaya Y, Shin TH, Miliaras N, Lee J, Oyama T, Mello CC. MEP-1 and a Homolog of the NURD Complex Component Mi-2 Act Together to Maintain Germline-Soma Distinctions in *C. elegans*. *Cell*. 2002.

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*. 2008;40(7):897–903.

Wingett SW, Andrews S. Fastq screen: A tool for multi-genome mapping and quality control [version 1; referees: 3 approved, 1 approved with reservations]. *F1000Research*. 2018;7.

Wolffe AP, Matzke MA. Epigenetics: Regulation through repression. *Science*. 1999;286(5439):481–486.

Wu X, Shi Z, Cui M, Han M, Ruvkun G. Repression of Germline RNAi Pathways in Somatic Cells by Retinoblastoma Pathway Chromatin Complexes. *PLoS Genet*. 2012 [accessed 2019 Jul 24];8(3):1002542.

Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications – writers that read. *EMBO reports*. 2015;16(11):1467–1481.

Acknowledgements:

Thank you to the everyone in the Katz lab: Dr. David Katz, Dr. Brandon Carpenter, Dr. Onur Birol, Dr. Teresa Lee, Dexter Myrick, Amanda Engstrom, Alyssa Scott, Juan Rodriguez, Sindy Chavez, Alicia Walker, Marcus Everett, and Semin Kang. Thank you everyone at Oglethorpe University: Dr. Karen Schmeichel, Dr. Lea Alford, Dr. Seema Shrikhande, Dr. Brent Runnels and Eli Arnold. Thank you to my mom Dr. Mirjana Brockett. Everyone of you helped me tremendously during this 2-year trek of my honor's thesis. Everything from talking science, taking the time to work with me, and provide moral support through encouragement and music enabled me to accomplish this task. I really appreciate everyone working so hard to make sure I succeed and finish my senior year with honors designation and something to hang my hat on. Lastly, thank you to the National Science Foundation for funding this project.



O G L E T H O R P E
U N I V E R S I T Y

Appendix:

Introduction: pg. 2-13

Methods: pg. 14-21

Results: pg. 22-29

Discussion: pg. 30-33

Citations: pg. 34-37

Acknowledgements: pg. 38

Appendix: pg. 39

Tables: 1-3 pg. 40-42

Figures: 1-11 pg. 43-53



O G L E T H O R P E
U N I V E R S I T Y

Trial	Genotype	(RNAi)	RNA Integrity (RIN Value)	[RNA] ng/ μ L
1	N2	(L4440)	8.30	45
		(<i>mep-1</i>)	7.20	42
	DM	(L4440)	7.20	19
		(<i>mep-1</i>)	7.50	17
2	N2	(L4440)	6.40	21
		(<i>mep-1</i>)	7.00	13
	DM	(L4440)	8.80	11
		(<i>mep-1</i>)	N/A	5

Table 1. Summary of total RNA extraction outcomes. Trials had no errors except *spr-5*; *met-2*; *mep-1* trial 2 as denoted by the “N/A” value under RNA integrity. This was because we did not gather enough RNA [$<5\text{ng}/\mu\text{L}$].



OGLETHORPE
UNIVERSITY

Samples	# tags	% tags after filtering	% aligned (once, >once)	# mapped reads
N2(L4440)Rep1R1	39,827,797	93.43	98.67 (94.14, 4.53)	36,715,632
N2(L4440)Rep1R2	39,827,797	93.43	98.67 (94.14, 4.53)	36,715,632
N2(mep-1)Rep1R1	30,946,252	92.60	98.0 (92.80, 5.20)	28,082,505
N2(mep-1)Rep1R2	30,946,252	92.60	98.0 (92.80, 5.20)	28,082,505
DM(L4440)Rep1R1	34,154,073	93.61	98.4 (93.66, 4.74)	31,459,888
DM(L4440)Rep1R2	34,154,073	93.61	98.4 (93.66, 4.74)	31,459,888
DM(mep-1)Rep1R1	34,416,598	93.56	97.64 (94.15, 3.49)	31,435,911
DM(mep-1)Rep1R2	34,416,598	93.56	97.64 (94.15, 3.49)	31,435,911
DM(L4440)Rep2R1	31,846,023	93.00	97.84 (91.98, 5.86)	28,975,351
DM(L4440)Rep2R2	31,846,023	93.00	97.84 (91.98, 5.86)	28,975,351
DM(mep-1)Rep2R1	36,326,627	93.45	98.16 (94.34, 3.82)	33,327,198
DM(mep-1)Rep2R2	36,326,627	93.45	98.16 (94.34, 3.82)	33,327,198
N2(L4440)Rep2R1	37,869,436	92.98	98.49 (94.06, 4.43)	34,678,054
N2(L4440)Rep2R2	37,869,436	92.98	98.49 (94.06, 4.43)	34,678,054
N2(mep-1)Rep2R1	35,052,087	92.52	98.3 (93.43, 4.87)	31,878,036
N2(mep-1)Rep2R2	35,052,087	92.52	98.3 (93.43, 4.87)	31,878,036
DM(L4440)Rep3R1	46,237,744	93.48	98.32 (94.15, 4.17)	42,497,957
DM(L4440)Rep3R2	46,237,744	93.48	98.32 (94.15, 4.17)	42,497,957
DM(mep-1)Rep3R1	36,740,795	93.55	98.36 (94.52, 3.84)	33,805,836
DM(mep-1)Rep3R2	36,740,795	93.55	98.36 (94.52, 3.84)	33,805,836
N2(L4440)Rep3R1	36,714,135	93.31	98.21(93.18, 5.03)	33,642,554
N2(L4440)Rep3R2	36,714,135	93.31	98.21(93.18, 5.03)	33,642,554
N2(mep-1)Rep3R1	42,669,380	91.31	98.01(93.53, 4.48)	38,187,725
N2(mep-1)Rep3R2	42,669,380	91.31	98.01(93.53, 4.48)	38,187,725

Table 2. Quality control analyses shows RNA sequencing was successful and able to map back to a known genome reference (ce10) <97.5% of the time. R1= forward strand, R2= reverse strand. Number of tags= total number of sequenced nucleotides. Percent tags after filtering = percent of nucleotide left after Trimmomatic. Percent aligned = number of tags that were able to match on ce10 genome using HISAT2. An alignment of <97.5% throughout the ce10 genome indicates high quality RNA and sequencing. Number of mapped reads= number of nucleotides able to be mapped back to ce10 genes using FeatureCounts.

UNIVERSITY

Genotype	Total # of differentially expressed genes	Up-regulated > 2 fold log ₂ FC	Down-regulated < 2 fold log ₂ FC
N2(<i>mep-1</i>)	77	2	0
DM(L4440)	4223	878	172
DM(<i>mep-1</i>)	7492	1790	1035

Table 3. Summary of differentially expressed genes in N2(*mep-1*), *spr-5*; *met-2*, and *spr-5*; *met-2*; *mep-1* mutants. (Significance cutoff of padjusted < 0.05). There is a significant increase in amplified genes as you increase in mutant genotype. All data normalized to wild-type. FC= fold change.



O G L E T H O R P E
 U N I V E R S I T Y

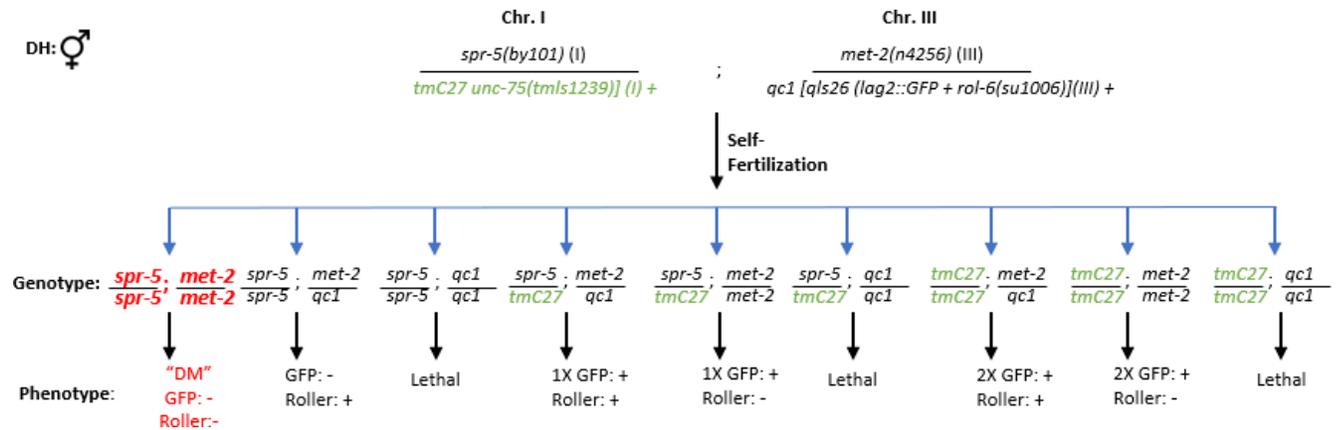


Figure 1. Selection Scheme for F1 DMs. Segregation patterns of *spr-5; met-2* dihybrids depicts every possible combination of these alleles and their expected phenotype. We only used homozygous animals (labeled in red "DM") for *spr-5* and *met-2* because they did not carry the wild-type balancer allele. They are indicated by having no GFP (GFP: -) and no roller phenotype (Roller: -). All others were recognized via GFP microscope, by eye, or were lethal.

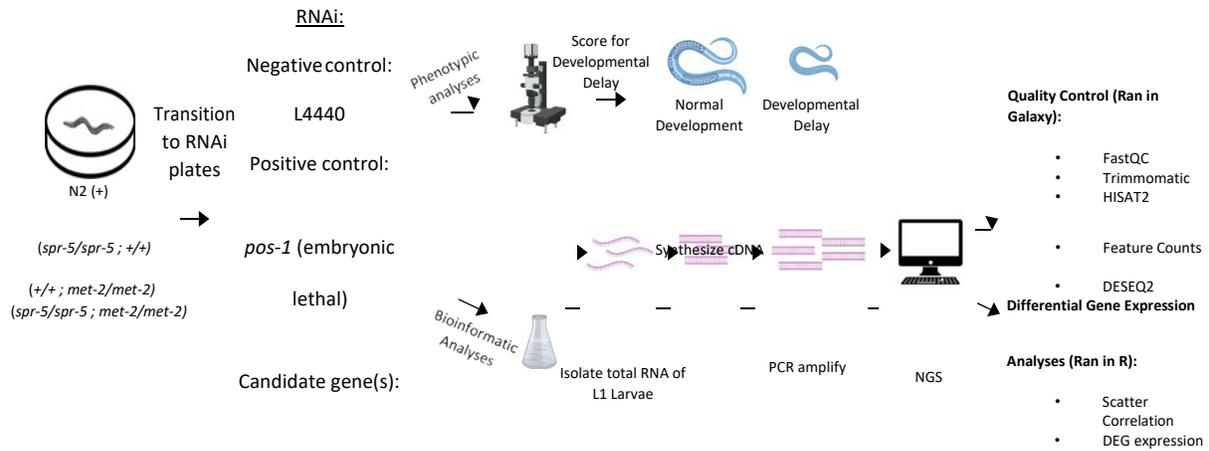
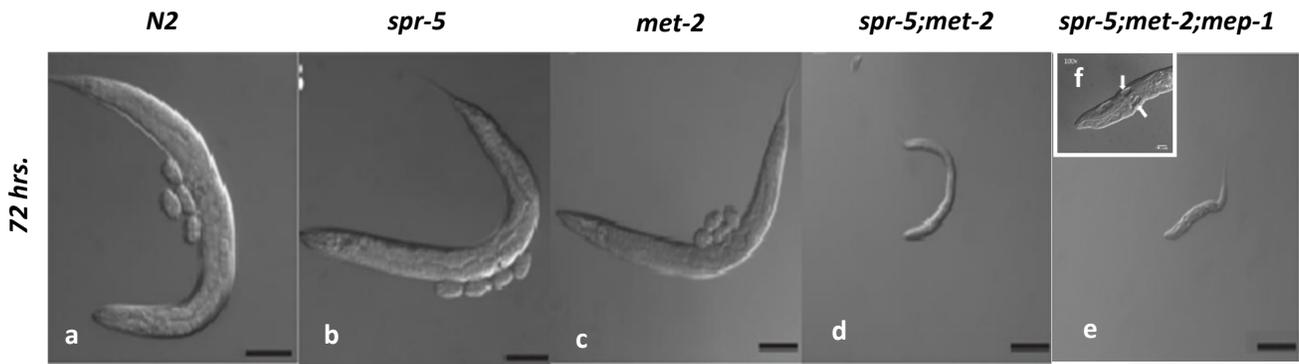


Figure 2. RNAi pipeline. The flow chart describes the sequence of events we used to isolate our progeny for L1 RNA-seq. First, each parental strain (N2, *spr-5*, *met-2*, *spr-5; met-2*) was selected for and subjected to each RNAi treatment. Then progeny of the parental strain were isolated and scored for development via light microscopy and/or DIC photography, followed by an RNA-seq analysis on L1 larvae.



OGLETHORPE
UNIVERSITY



g

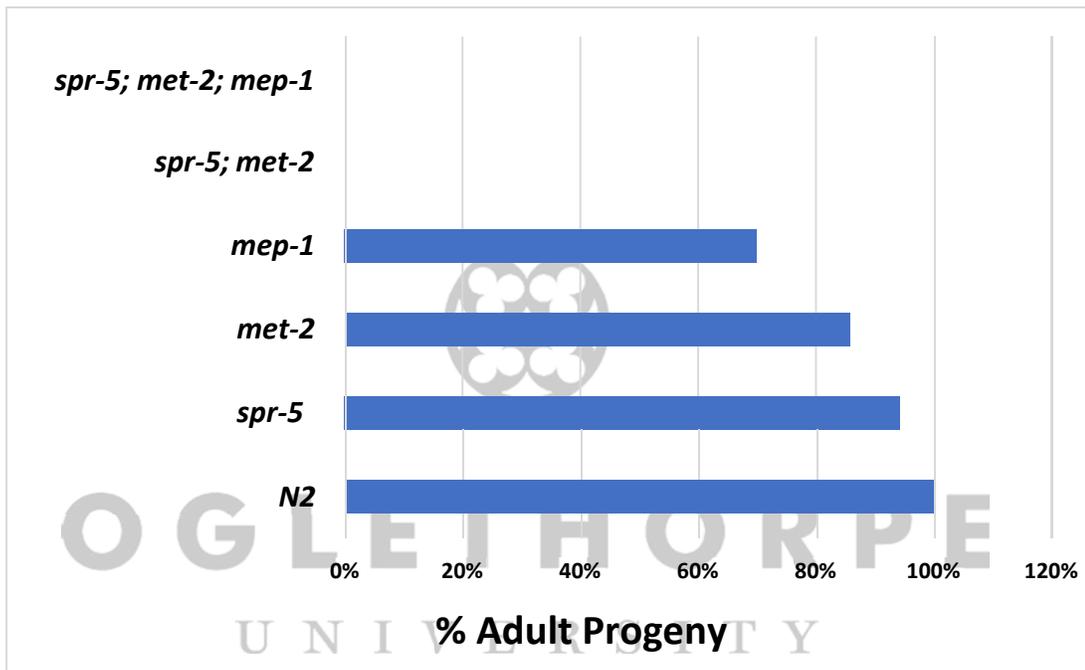


Figure 3a-g. Loss of *mep-1* in *spr-5; met-2* progeny, exacerbates developmental delay phenotype. The results of the phenotypic analyses show that after 72 hours post fertilization, both *spr-5* and *met-2* single mutants exhibit little developmental defects (Figure 3b, c). However, *spr-5; met-2* double mutants experience a severe developmental delay at the larval stage 2 (Figure 3d). Similarly, *spr-5; met-2; mep-1* triple mutants exhibit a severe developmental arrest (Figure 3e) along with structural defects in the cuticle towards the head not seen in the *spr-5; met-2* double mutants (Figure 3f). Percentage of *N2*, *spr-5*, *met-2*, *spr-5; met-2*, and *spr-5; met-2; mep-1* progeny that reached adult stage (% adult progeny) by 72 hours post synchronized lay (Figure 3g). All single mutants retain >75% of normal development. Introduction of two mutations into either one of the singly mutant backgrounds results in a dramatic decrease in progeny. (Data not shown for *spr-5; mep-1* and *met-2; mep-1* double mutants) Data adapted from Carpenter et al 2020 and Chavez 2019.

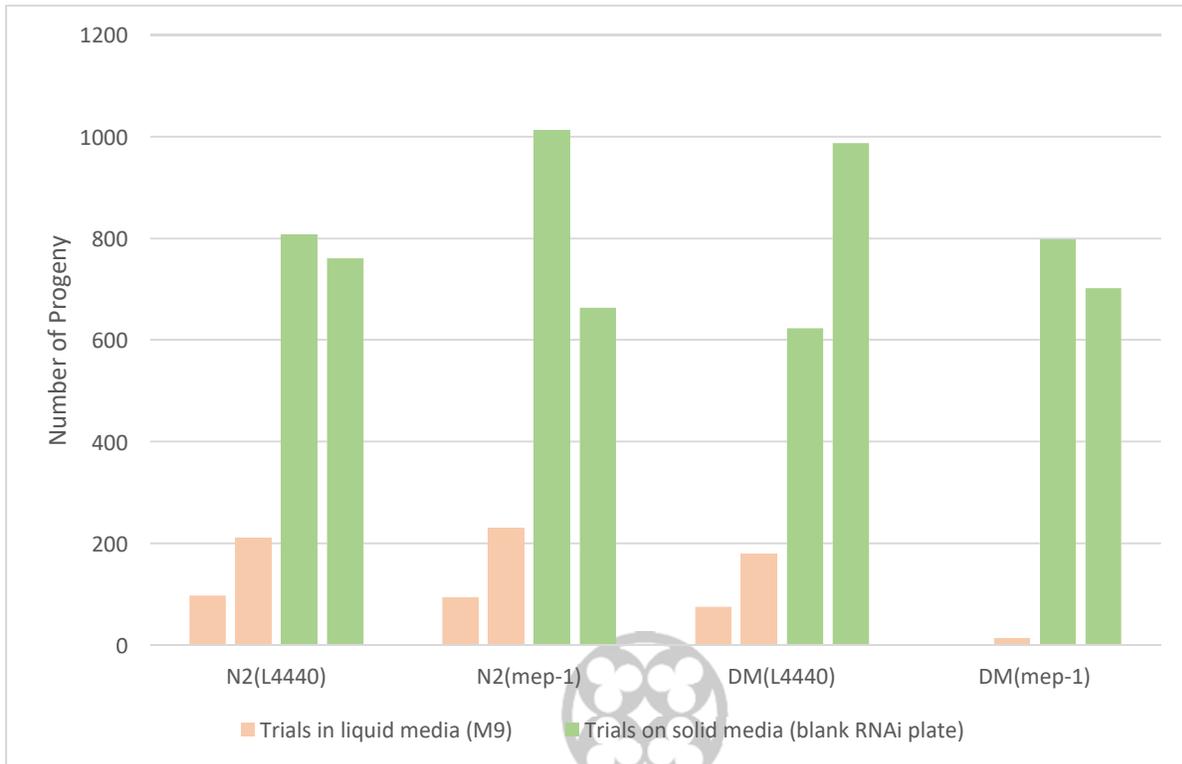
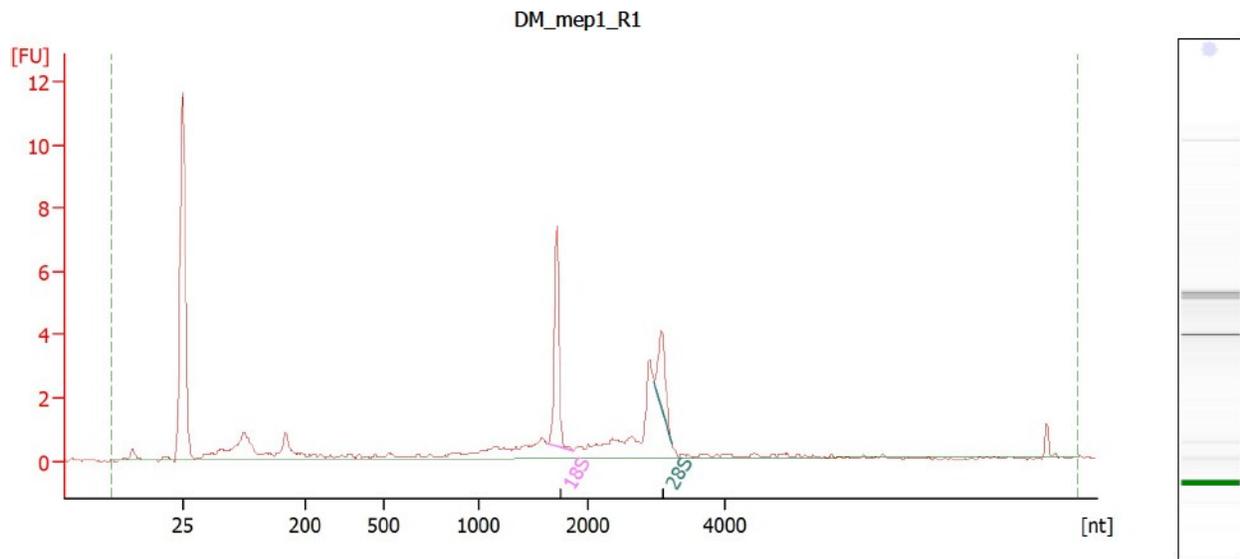


Figure 4. Revamped protocol allows for increased amount of progeny per experiment. Trials using M9 buffer for parental generation to lay F1 progeny are in orange. Trials using empty RNAi plates for parental generation to lay F1 progeny are in green. Upon optimization, experiments using empty RNAi plates reveal that this increases brood sizes substantially while keeping the integrity of the experimental controls.



Overall Results for sample 5 : DM_mep1_R1

RNA Area:	32.4	RNA Integrity Number (RIN):	7.5 (B.02.08)
RNA Concentration:	17 ng/μl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.4	Result Flagging Label:	RIN: 7.50

Figure 5. Example of RNA integrity data indicating *spr-5*; *met-2*; *mep-1* sample had undergone proper extraction without RNA degradation. The presence of peaks denoted as 18s and 28s indicates that RNA is intact and was not exposed to RNases during the extraction process. Total RNA concentration for the sample is shown by RNA concentration: 17ng/μL. The RNA integrity number (RIN) is calculated using the total concentration and height of the subunit peaks. An RIN value of >7 typically denotes a good score and quality RNA.

UNIVERSITY

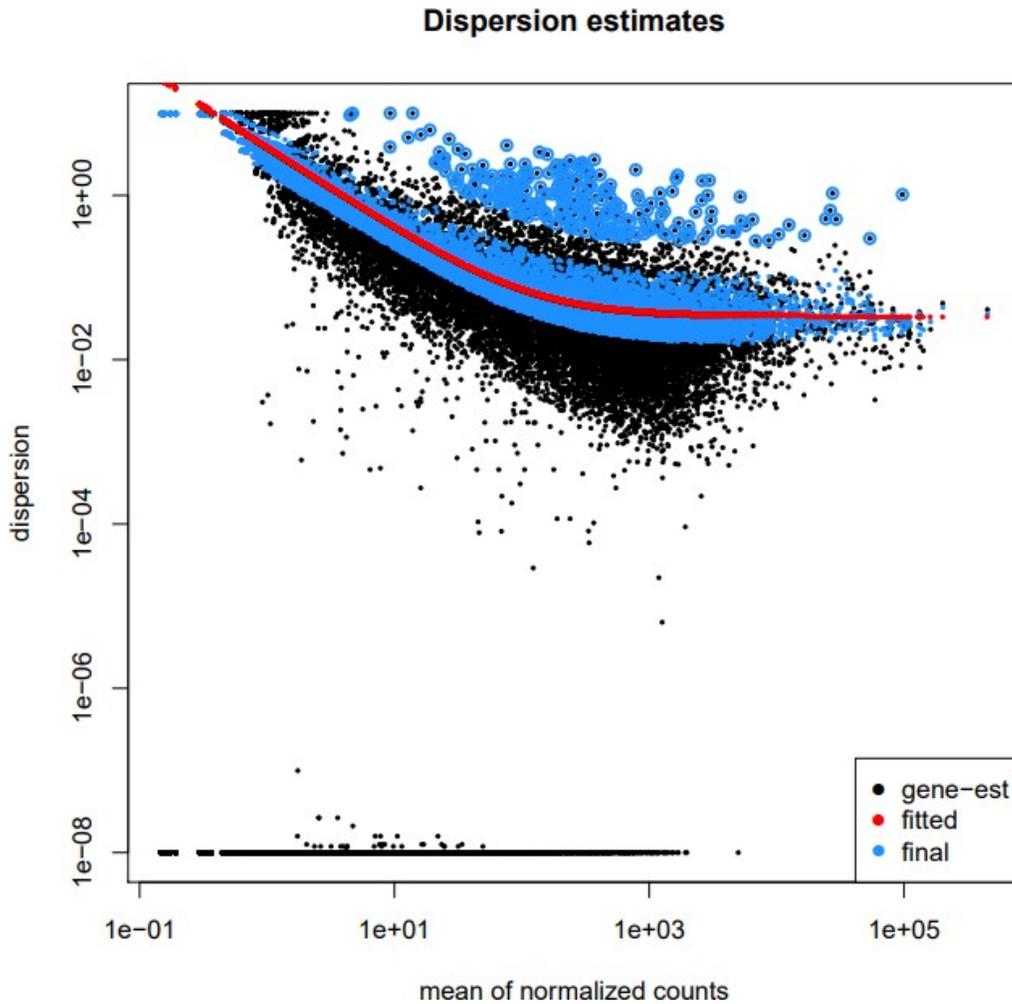


Figure 6. Dispersion plot of mapped reads in DESeq2 shows correlation between the mapped reads and amount of amplification. After Feature counts gathers total mapped reads, DESeq2 conducts various correlations analyses to confirm that the mapped reads are either from the same organism and that they are being affected in the same way. The expectation is to see the genes (black dots) fit closely to the fitted red line indicating they are changing in the same direction. Example from *spr-5*; *met-2*; *mep-1* mutant RNA.

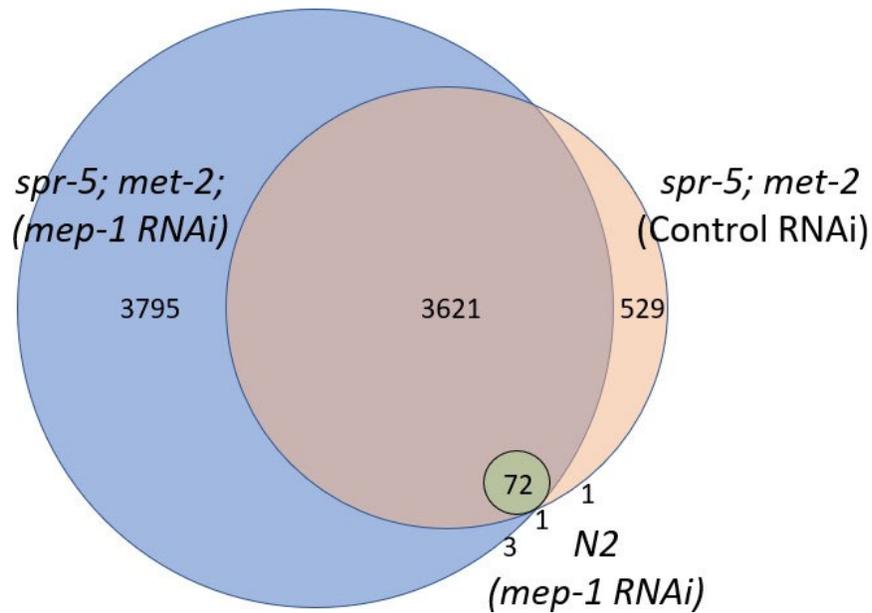


Figure 7. Overlap of differentially expressed genes between N2(*mep-1*), *spr-5; met-2*, and *spr-5; met-2; mep-1*. The Venn diagram shows a significant (padjusted <.05) overlap of differentially expressed genes between *spr-5; met-2* and *spr-5; met-2; mep-1* mutants. The numbers correspond to the number of genes differentially expressed in that specific mutant. The numbers inside the overlap represent the number of genes differentially expressed in both mutants compared to wild type. All data normalized to wild type.

UNIVERSITY

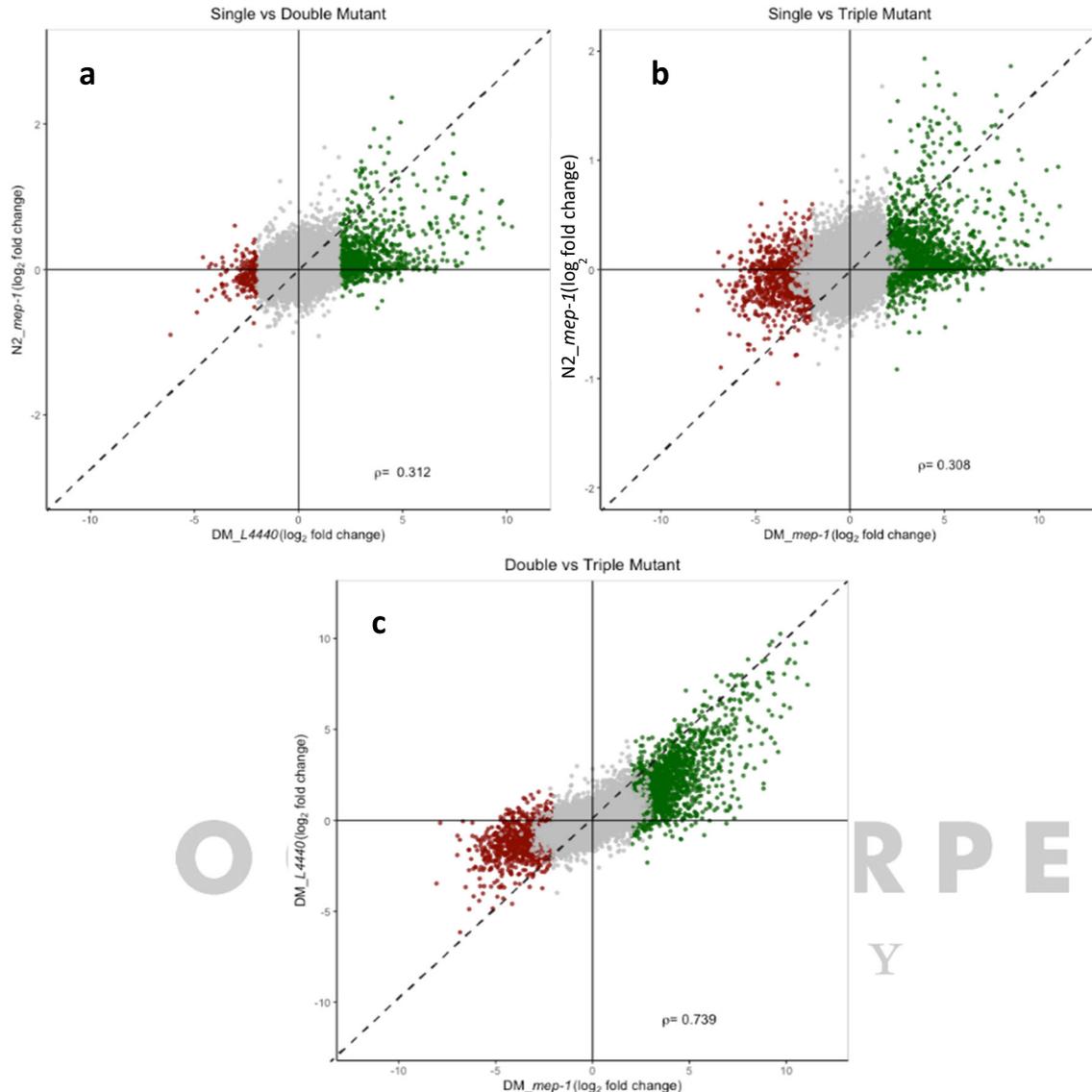


Figure 8a-c. Scatter plot correlation between N2(*mep-1*) vs. *spr-5*; *met-2*, N2(*mep-1*) vs. *spr-5*; *met-2*; *mep-1*, and *spr-5*; *met-2* vs. *spr-5*; *met-2*; *mep-1* demonstrates the same differentially expressed genes become more amplified in more mutant backgrounds. The scatter correlation plot demonstrates the difference in expression between the genes in the least mutant genotype on y-axis and the genes in the most mutant genotype on x-axis. Green represents genes that are significantly upregulated, red represents genes that are significantly downregulated, and gray represents genes that are not significantly expressed. The 1:1 line indicates genes that have no change in levels of expression between genotypes. Looking at N2(*mep-1*) single mutants vs. *spr-5*; *met-2* double mutants, the data demonstrates that the same genes in the N2(*mep-1*) single mutant are expressed at higher levels. Comparing this to N2(*mep-1*) vs. *spr-5*; *met-2*; *mep-1*, there is a noticeable shift in genes (up and down regulated) from the N2(*mep-1*) single mutant towards higher expression profiles in the *spr-5*; *met-2*; *mep-1* triple mutants. Lastly, when comparing *spr-5*; *met-2* vs. *spr-5*; *met-2*; *mep-1* differentially expressed genes, there is a similar trend where the same genes in the *spr-5*; *met-2* double mutants become even more amplified in the *spr-5*; *met-2*; *mep-1* triple mutants.

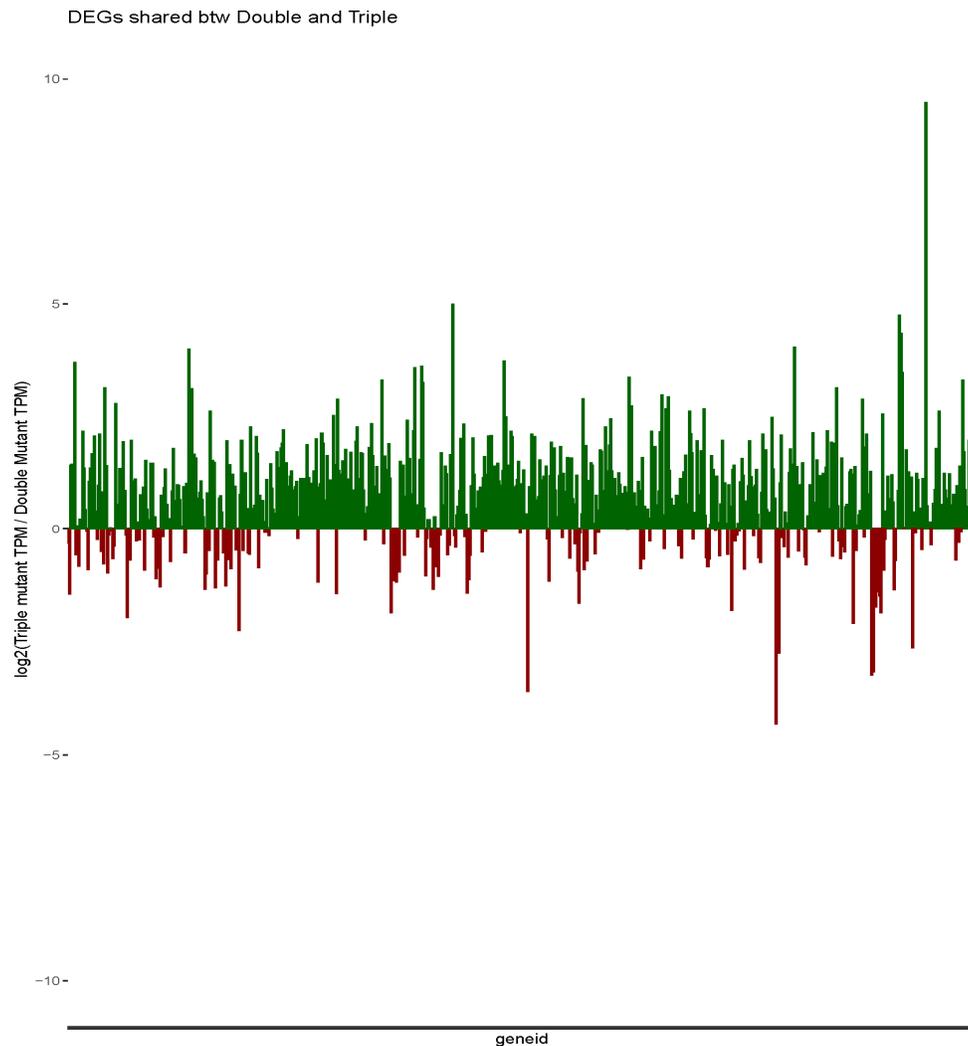


Figure 9. The same genes that are differentially expressed in *spr-5; met-2* mutants are even further expressed in *spr-5; met-2; mep-1* mutants. The *spr-5; met-2* genes that were differentially expressed were set to 0 in order to show the significant difference in amplification between *spr-5; met-2* mutants and *spr-5; met-2; mep-1* mutants. The genes significantly up-regulated are in green and the genes significantly down-regulated are in red. The amount of expression (\log_2 fold) is in units of transcripts per million (TPM). All genes in the *spr-5; met-2* mutants were previously normalized to wild type and then set to 0 to create a baseline for the *spr-5; met-2; mep-1* differentially expressed genes.

RNAi: L4440

L4440-*mep-1*

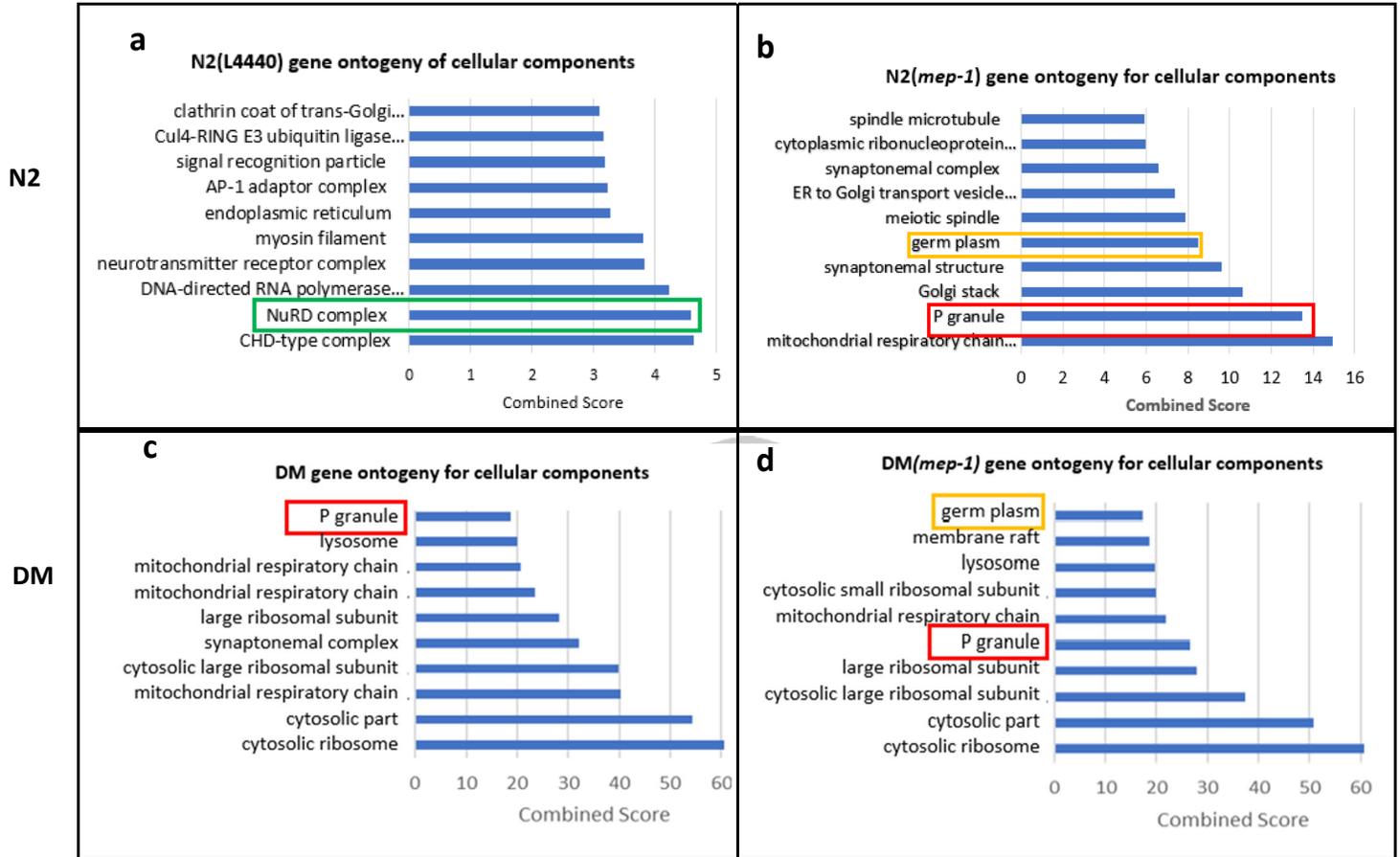


Figure 10a-d. Gene ontology reports of cellular components of each genotype reveal germline gene misexpression in all mutants and an exacerbated effect the more mutant the phenotype. In wild type animals, L1 larvae are expressing genes associated with somatic cell development such as: endoplasmic reticulum development and interestingly enough, genes associated with germline gene repression (NuRD complex genes, Green). In N2(*mep-1*) single mutants, germline genes such as genes associated with the formation of germ plasm (yellow) and p granules (red) begin to express themselves in the soma of the L1 animal, however in low quantities. In the *spr-5; met-2* double mutants, those same p-granule genes are being further misexpressed indicated by the scale change of combined score from 14 in N2(*mep-1*) single mutant, to nearly 20 in the *spr-5; met-2* double mutant. Lastly, in the *spr-5; met-2; mep-1* triple mutants, germ plasm and p granule are even further shifted up than in previous mutants indicating that those genes are being even further expressed.

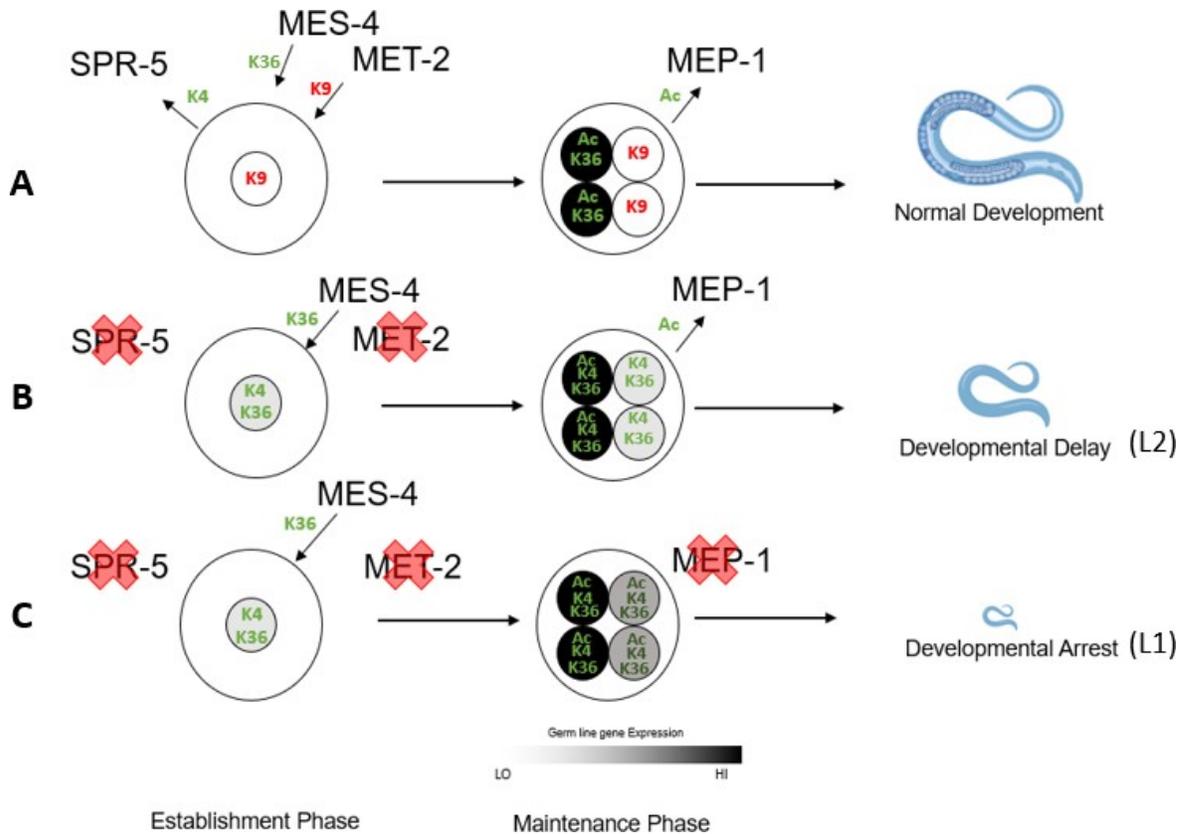


Figure 11. Diagram of SPR-5, MET-2, MES-4 and MEP-1 interactions that dictate germline vs somatic cell distinction. At fertilization, SPR-5 and MET-2 turn off gametogenesis genes in the soma that were previously left on from the parent to form the gametes. At the same time, MES-4 bookmarks those same genes with H3K36me3 in order to keep the germline genes poised in cells destined to be germ cells. MEP-1 works collaboratively with SPR-5 and MET-2 to further reinforce the action of: *turning off germline genes in the soma*. This action is acquired through deacetylation of the same gene targets as SPR-5 and MET-2 to further repress any germline specific genes in the soma. Organisms deficient in SPR-5 and MET-2 experience a severe developmental delay due to the misexpression of germline genes from ectopic H3K36me3 from MES-4. However, the delay is not permanent presumably because of MEP-1 and other maintenance phase chromatin repressors still shutting down germline genes through deacetylation. Organisms deficient in SPR-5, MET-2, and MEP-1 result in a permanent arrest at the L1 stage. Since many levels of epigenetic repressive complexes are eliminated from the organism, MES-4 can have increased accessibility to chromatin throughout many cell divisions and mark more ectopic H3K36me3 on greater sets of germline genes. The result of this leaves the animals fully arrested without any recovery. Primordial germ cell = black, somatic cell = white, grey cell = combination of the two, dark grey = combination but more germ cell than soma.