Moving Forward Genetics Forward With a New RNA-Seq Based Approach

February 18, 2013

GMR Deyter

Forward genetic screens are a method used to identify genes which are responsible for a particular phenotype. They have been used to identify many molecular components of the biochemical pathways that develop and maintain life. However, the identification of RNA interface (RNAi), a biological process by which RNA molecules inhibit gene expression, and other powerful genetic knock-down strategies, has made reverse genetic screens extremely popular. Reverse genetic screens start by disrupting a known gene, and then analyze the resultant phenotypes. Many forward genetic screens have been less fruitful given the difficulty in mapping causative mutations of a given phenotype. This is especially true in vertebrate models, where whole genome sequencing (popular with invertebrate screens) can be costly due to large genome sizes. The Moens's Lab (Basic Sciences Division) recently developed a novel technique based on RNA-seq to provide a relatively cost effective way of identifying mutations from forward genetic screens.

Under Dr. Moens's guidance, Miller *et al.* devised an RNA-seq-based bulk segregant analysis (BSA) approach to identify the candidate mutations. BSA identifies regions of the genome that are linked to a mutation in a group of mutant animals. The researchers reasoned that next generation sequencing of RNA isolated from mutants would uncover single nucleotide polymorphisms (SNPs) that co-segregate with the mutant phenotype and that would help to map mutations. Moreover, changes in mRNA transcript levels or splicing events would be revealed by the sequencing data and might therefore give the researcher a head start in discovering how the mutation might affect protein stability or function.

As a proof of principle, Miller and colleagues used four already-established zebrafish mutants to test whether RNA-seq would effectively identify the genomic locus of the mutations. After crossing heterozygous mutants (to obtain wild-type siblings from the cross) and scoring for the phenotype of interest, RNA from wild-type and mutant embryo pools was extracted, and cDNA was prepared from the isolated mRNA to generate the sequencing libraries (see figure). By barcoding each pool with a unique sequence tag, a total of six libraries (three sibling/mutant pairs) were added to each lane of an Illumina HiSeq 2000 machine to generate an average of 43 million 50-base pair paired-end reads per sample.

The reads from each sibling/mutant pair were aligned to the zebrafish genome using the TopHat/Bowtie program, and SNPs within the wild-type sequence were identified that could serve as markers to test for linkage within the mutant sequences. 40,000 high-confidence markers (SNPs) per experiment were identified that were used to determine if a particular genomic region was linked to the mutation of interest. For instance, the SNP marker frequency at or proximal to the mutation will be 1 (all alleles are the same), with a decrease in frequency the farther away the SNP resides on the chromosome from the mutation. The researchers confirmed that their RNA-seq protocol successfully

identified the four known genetic lesions of their mutants. Increasing the number of embryos sequenced from 20 to 80 helped to decrease the linkage area, indicating that sequencing a moderate number of embryos can accurately pinpoint the genomic region of the responsible mutation. Moreover, the authors gathered that their RNA-seq approach could identify a reasonable number of SNP changes that could represent mutations of interest when mapping *unknown* mutants from genetic screens.

Altogether, the development of an RNA-seq based mapping approach to identify candidate mutations from genetic screens has many advantages. RNA-seq is relatively inexpensive and effectively limits the genomic landscape sequenced since only expressed genes would be sequenced. Potential changes in RNA splicing due to mutations would be apparent within the sequenced datasets, as would changes in the expression level of genes linked to the mutation of interest. The authors have also developed an *in silico* bioinformatics pipeline (www.RNAmapper.org) to aid other researchers in their genetic mapping endeavors. The novel RNA-seq method developed by Miller and colleagues will undoubtedly hasten the final steps in forward genetic screens to reveal mutations in genes that affect nearly every aspect of life.

Miller AC, Obholzer ND, Shah AN, Megason SG, Moens CB. 2013. RNA-seq-based mapping and candidate identification of mutations from forward genetic screens. Genome Res. doi: 10.1101/gr.147322.112
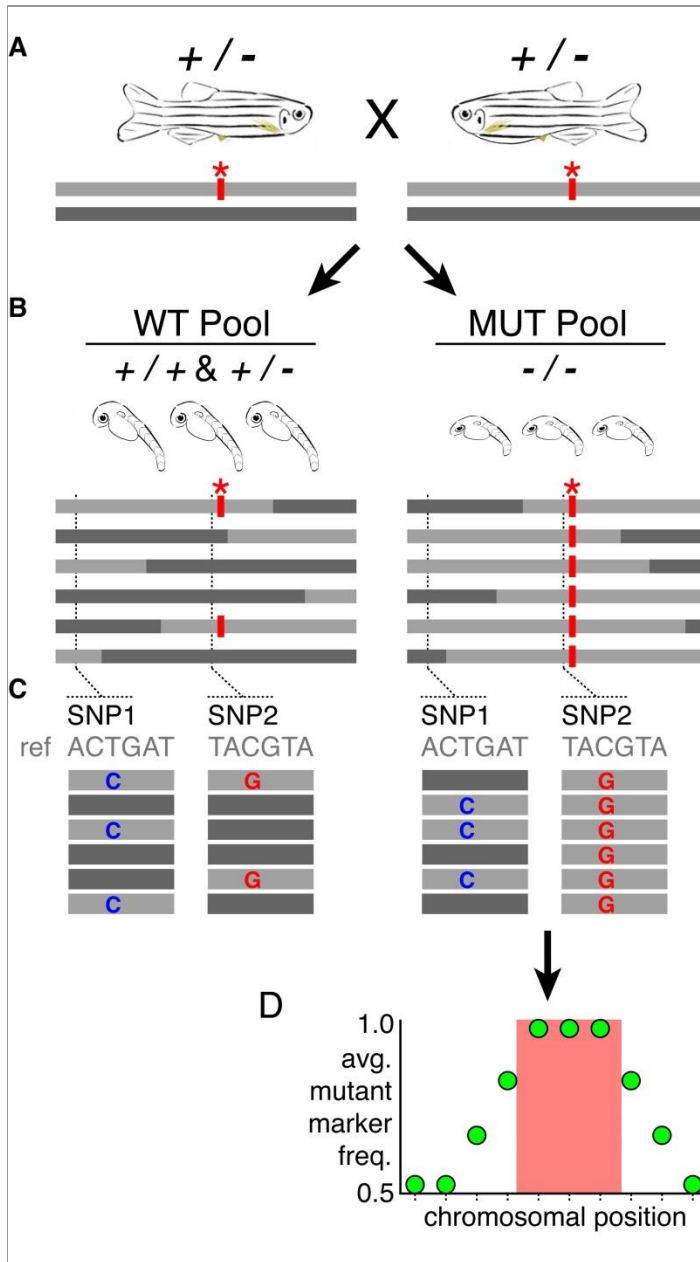
Image obtained from the manuscript

Figure: Diagram of the RNA-seq-based approach. A) Heterozygous animals (+/-) that have a mutation (red asterisk; grey bar represents chromosome linked to the mutation) are crossed to obtain wild-type and mutant embryo pools. B) The mutant pool will be homozygous for the mutation and also for SNPs proximal to the mutation (i.e. SNP2) but will be heterozygous at other SNPs (i.e. SNP1) due to meiotic recombination. C) The wild-type pool identifies SNP1 and SNP2, two polymorphisms useful for mapping. The sequence data from the mutants is then interrogated at the marker positions to identify homozygosity, indicating the identification of the candidate mutation. D) Plotting mutant marker frequency against chromosomal position reveals the region of linkage (red box).