Scientific premise: The significance of this application is based on the scientific premise that there is a need for reducing the number of candidate driver processes so that research efforts can be focused on those more likely to result in effective therapeutics. A good review of the diversity of processes affected from a molecular perspective is (6). The most recent publication concerning the many changes in HD found by RNA sequencing of cortical samples from postmortem HD samples showed half of the transcriptome is differentially expressed in HD cortex which produced several overrepresented functional groups by ontology analysis (9). To our knowledge, our preliminary analyses and the proposed work are the first to address the need for more powerful screening methods using the graded pathology of an allelic series (13).

Authentication of Key Biological Resources. Since the removal of genetic background effects is an important feature of the knock-in mouse lines we have taken the following steps to insure that the lines used in this work are congenic. First, we have breed back to C57BL/6J mice for 12 generations. Second, we analyzed our RNA sequencing results to determine the extent of 129 background from the ES cells used to insert the mutation vs. C57BL/6J, the recipient strain. There are over 9000 expressed SNPs between these strains and our RNA sequencing results found that all sequences matched the C57BL/6J background. Furthermore, of 25 million 50 base pair reads in each of 3 mutant lines of the series we only found 3 sequence differences from C57BL/6 (13) none of which represented a known SNP vs. 129. Since all other strains of the series were derived from an originally congenic HDQ150 line, the sequencing of this line authenticates the entire series.

Consideration of Relevant Biological Variables: The ordering of members of the allelic series depends on their phenotypic severity being driven by the same molecular processes. This assumption is a reasonable one given the phenotypic similarities between the lines. Nevertheless, there is one known molecular feature that differs qualitatively. Striatal aggregates of Htt protein are nuclear for repeats with 200 or fewer CAGs. The HDQ250 line has many aggregates that are extranuclear and in the HDQ315/+ line extranuclear aggregates predominate (See Fig. 5 and (13, 24, 29-31)). Our prior results show HD315/+ mice have a more robust phenotype than HDQ150/+ mice suggesting that nuclear location of aggregates does not associate with phenotypic severity across the series. Nevertheless, the intracellular location of aggregates may be important to certain RNA levels given the importance of the nuclear compartment in mRNA expression. Therefore we will perform a separate analysis where lines with aggregates exhibiting nuclear and extra-nuclear are segregated. Each transcript will be annotated as to whether its levels change with increasing nuclear aggregate load, increasing extranuclear aggregate load or both. Transcripts with level corresponding with nuclear aggregate loads will be classified as correlates and discorrelates with respect to phenotypic severity using only the lines in the series with predominantly nuclear aggregates (200 CAGs and fewer) while those transcripts with levels corresponding to extranuclear load will be classified bases on the phenoptypic severity of only those with predominantly extranuclear aggregates (250 CAGs and greater). This analysis will be reported in our GEO files for all to see and compared to analysis across the entire series to determine whether nuclear location of aggregates influences each transcripts level. Those influenced will be annotated as such.

Scientific Rigor and Transparency The use of 20 biological replicates for RNA sequencing analysis may seem like too many given the cost of obtaining this data. This analysis is, however, not a simple mutant vs. wild type comparison. Differentiation between mutants is central to us achieving our aims therefore, standard power analysis between mutant lines (shown in application) was used to achieve the rather high n that will be required to make rigorous conclusions from our data. As far as transparency, all data will be provided on the GEO database including the behavioral output measures linked to transcript level data for individual mice. Furthermore, all of our mouse lines are already available from Jackson labs, for anyone wishing to replicate this work.