White Paper

Biologically constrained behavioral genetics research

Elizabeth A. D. Hammock
Department of Pediatrics
Vanderbilt University School of Medicine
1115 Light Hall
Nashville, TN 37232
liz.hammock@vanderbilt.edu

A bottom-up perspective on social behavior

From a biologist's perspective, social behavior includes any behavior that involves at least two actors. By this definition, social behavior can include aggregation in slime molds, the colony structure of the eusocial insects, or the coordinated efforts of humans across vast distances to successfully land on the moon. The diversity of this range of behavior shares one driving force: natural selection. While natural selection acts at the level of phenotype (e.g., morphology, metabolism, behavior) the ultimate unit of natural selection is the gene contained in DNA—the object of inheritance. The relationship between DNA and social behavior is uncovered in the field of *sociogenomics*, defined as the mechanistic study of genes, gene products, and gene × gene interaction networks supporting emergent social behaviors.[1,2,3]

From a neuroscientist's perspective, the brain is an experience-expectant organ that serves to maintain homeostatic balance. The main behavioral tasks that the brain has been selected to perform are the "four F's": feeding, fleeing, fighting and mating. Mammalian brains have added layers and layers of control to these behaviors. Extra layers of control correlate with larger brains, requiring longer postnatal development before reaching full maturity. The extra experience-expectant developmental time allows humans to gain skills to manage and manipulate their environment, including the social environment. There is significant variation in this ability across individuals as well as across time within a given individual. This variability must come from individual differences in brain structure/function, which in turn derives from genes, environmental influences, and their interaction throughout development and into maturity.

doi: 10.2990/30_2_93

Really, what is a gene?

Historically, textbooks have defined a gene as a linear sequence of DNA that is translated into protein by specialized cell machinery. Proteins were awarded the status of intermediary between DNA and phenotype. In recent decades, however, this view has been clarified. It turns out that "genes" are really laborious to define with precision and are of somewhat indeterminate size, and proteins are not the only product of DNA with functional properties. The hardest part of defining a gene is determining which parts of the genome carry instructions for when and where a gene product should be made. Should a protein be present only in neurons that produce glutamate or neurons that produce some other neurotransmitter? Should it be present throughout the lifespan of the individual, or only during development, or only after mating—or in a fight-or-flight moment?

Some gene control regions occur in the linear DNA sequence immediately before the protein coding region starts. Some control regions occur immediately after the coding sequence has ended. And there are even control regions for a given gene that may exist thousands of base pairs away from the protein-coding region of the gene, and these may sometimes be intermingled with other protein coding regions for separate genes. A protein or other product of a DNA sequence is nothing without the

https://doi.org/10.2990/30_2_93 Published online by Cambridge University Press

instructions for where and when it should be expressed. As an illustration, there are very few differences in protein coding regions between chimpanzees and humans; the majority of the genomic differences reside in the non-coding gene control sequences.

How do individual genomes differ?

Gene variation among individuals contributes to variation in brain structure and function, which contributes to differences in social behavior. There are several kinds of individual variation in genomes, including varition which affects individual nucleotides, sequence length, or its relative position in the genome. These include single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), inversions, insertions/deletions (INDELs), and copy number variation (CNV). These changes in the sequence of DNA can occur within coding regions of genes and can alter protein activity, or they may occur in non-coding regions of DNA where they do not alter the transcribed portion of a gene but instead alter the quantity and location of a protein or other gene product.

How do we assess genome differences? Long sequence reads allowed us to piece together the draft genome. These long reads employ labor intensive methods and are cost-prohibitive. Newer techniques involving automated sequencing by synthesis are high throughput allowing for very short sequence reads, but these techniques force the genome into a single size mold and are not technically capable of discovering certain kinds of gene variation, including STRs and inversions. There are several new technologies emerging on the horizon that can make up for the gaps in methodology. One such method involves threading single DNA molecules through a nanopore detector that immediately reads the sequence.[4] When applied in parallel, multiple long DNA molecules can be read simultaneously. Theoretically, this permits determination of all kinds of gene variation. The newest arrival on the horizon of whole genome analysis involves high throughput ascertainment of the significant inter-individual structural variation in the human genome,[5] with promised structural analysis of individual genomes for less than $1,000 in under an hour.

Why is genetic methodology so important? Not all technologies are capable of recognizing all kinds of gene variation. There is growing frustration in medical and psychiatric genetics for the disappointing strength of findings in gene association studies as researchers continue to look for "the missing heritability."[6] The causal gene in single-gene disorders such as Huntington's disease, Rett syndrome, and Fragile X, has been identified. In contrast, we are still searching for gene variation that will robustly explain risk for disorders with high heritability, including autism and schizophrenia. In genome-wide association studies, SNPs have been the main target for investigation. This approach has yielded some insights into risk alleles that may contribute to these disorders.[7,8] Both common alleles and rare mutations have been studied and both offer some explanation. Promising new insights have come from focusing not on SNPs but on other kinds of genomic variation, including structural variation in the genome.[9,10,11,12] To make steady progress in understanding how genetic variation leads to behavioral variation in health and disease, *all* kinds of genomic variation should be considered, both within and outside of coding regions.

What is known?

To offer an indication of what we know about gene variation and social behavior, let us look at a few examples from the literature that relate each kind of genetic variation to individual or species differences in social behavior. Here, I will consider SNPs, then STRs and finally CNVs and structural variation. The examples reviewed are selected because of the strong correlation between a given gene variant and behavior, and because they have demonstrated mechanistic evidence at the level of the nervous system. Such evidence adds weight to the gene:behavior relationship. Because of this, these examples are limited to animal studies and gene studies in human disease.

Single Nucleotide Polymorphism (SNP) and social behavior. The roundworm *C.elegans* is a well-known tractable model system in biology. While the social behavior of the roundworm is severely limited compared to that of humans, or vertebrates in general, we have been able to learn about mechanisms of genetic variation related to social behavior variation from the roundworm. Roundworms are a diverse taxa. Some strains of roundworms feed in social groups while others

https://doi.org/10.2990/30_2_93 Published online by Cambridge University Press

are solitary feeders. Variation in a single nucleotide in the coding region for a g-protein coupled receptor that looks like a mammalian neuropeptide y receptor (*npr-1*) can influence this behavioral tendency.[13,14] Roundworms feed on bacteria, which can contain noxious stimuli. Noxious stimuli can cause worms to aggregate, where they may benefit from cooperative secretion of enzymes to inactivate bacterial toxins, for example. One version of the *npr-1* strongly inhibits aggregation in response to signals from bacteria. The other version of *npr-1* does not inhibit aggregation. Further, the probability of social aggregation is directly related to such environmental influences as population density and the food satiety of the animals. The molecular and cellular players in detecting population density and satiety create an added layer of control.

This example is not only informative in the details of gene and social behavioral variation but it also illustrates the molecular mechanisms of cooperation and competition that subserve social behavior (e.g., whether or not the worm should aggregate and cooperate to feed or remain in isolation).

Short Tandem Repeat (STR) and social behavior. STRs are more difficult to measure than SNPs and they incur statistical costs during analysis because of the high levels of polymorphism. There are multiple examples of STR variation playing a role in social behavior phenotypes. The vasopressin 1a receptor (*Avpr1a*) promoter region in voles has become a popular example. A long history of behavioral neuroscience research on this receptor and its ligand, vasopressin, implicated this system in mediating prosocial attachment related behaviors in adult voles. In a series of comparative genetic studies, the *Avpr1a* upstream regulatory region was compared across four closely related species of voles. The length of an STR appeared to correlate with the species' social structure.

This simple finding led to a testable hypothesis that intraspecific variation in repeat length might be associated with intraspecific variation in social behavior. A breeding scheme that selected for longer STR length resulted in offspring that had a higher probability of forming a mate preference, increased social approach behavior, and altered brain expression of the *Avpr1a* gene product.[15] This one example represents a minute portion of the possibility of the thousands of potentially functional STR sites in the human genome. Bioinformatic metanalysis of the genome indicates that many genes involved in neurodevelopment are enriched for STRs. This is an understudied area that deserves more attention.

Developmental biologists will need to determine the developmental impact of relevant STRs and they would do well to accept the aid of statisticians who can help solve problems related to statistical power when there are so many allele possibilities.

Copy Number Variation and Structural Variation and social behavior. There are several disorders of brain development where structural variation and CNVs are implicated or causal. These include autism and schizophrenia, Williams syndrome, and Prader-Willi/Angelman syndrome. There are no examples yet of CNV or structural variation within the range of typical behavior in humans. In humans, mice and rats, there is evidence that normal copy number variation across strains contributes to variation in gene expression.[16,17,18] It is not a stretch to consider a similar mechanism for individual differences in genes in the brain contributing to normal variation in behavioral traits. This is a promising area of future research.

What should we know about genes and social behavior?

As we look at the near term (i.e., by 2020) for research areas in which to invest precious research dollars, there are transformative ideas and curiosity-driven research questions that come to mind.

Near term research needs in new or understudied areas:

1) *Genome variation and behavioral variation in infancy and childhood*. In an effort to understand the strength of gene:behavior variation, it is reasonable to hypothesize that the earliest social orienting behaviors might have robust genomic correlates and that such gene effects are diluted over time due to environmental variance or gene × gene interaction as more brain systems and layers of control are functionally integrated.
2) *Better understanding of sequence and structural variation in the human genome in healthy individuals*. This is a pressing topic in the face of the "missing heritability" question. Research is needed in coding and non-coding variation.
3) *Assessment of CNV/structural genome variation and typical social behavior variation*. This

https://doi.org/10.2990/30_2_93 Published online by Cambridge University Press

includes advances in statistical and computational approaches to determine significant relationships.

4) *Gene association studies are best understood in functional biological contexts.* In particular, gene association studies are best informed by an understanding of gene function in the biological system and when and where is it expressed. This is a critical need in human development. Recent collaborative efforts by several entities have created two resources to provide the first glimpse at answers to these important questions: Brainspan Atlas of the Developing Human Brain (www.brainspan.org)[19] and the Human Brain Transcriptome (hbatlas.org).[20,21] These expression maps let us view, for the first time, the developmental trajectories of gene products in a variety of areas within the human brain. Ideally, we would also have cell-type specific information, which is technically feasible, but at very high cost.

5) *Epigenetic modifications by experience and parental inheritance.* There is a need to increase our understanding of the mechanisms of how the environment and the genome interact to impact the expression of an individual's genetic endowment, both within an individual's lifetime and across generations.

Curiosity driven research areas:

1) The neurobiology of in-groups and out-groups
   a. Social phobia
   b. Political unrest
   c. Population genomics of jingoistic policies?

2) Social matching (genetic and behavioral typing)
   a. Mate finding
   b. Job matching
   c. Parent-child/caregiver-child matching in foster care?
   d. More efficient rehabilitation methods in criminal justice system?

3) Social ontogeny
   a. Post-natal social programming by experience-dependent epigenetics
   b. Are there other experience-dependent changes to the genome?
   c. What are the developmental sensitive periods?
   d. What are neurobiological effector molecules?
   e. Is it possible to re-open sensitive period plasticity for therapy?

4) How does genetic sex or gender identity interact with the genome?

5) Comparative genomics among social and non-social congeners
   a. Voles
   b. Primates
   c. Domestic dogs

6) How do social and stress systems in the brain interact, and how does this relate to gene × gene interactions?

7) What potential impact can economic and/or social policy have on human genetics?
   a. With cultural bias for males, is there genetic selection in China after one child rule? This could serve as a bellwether for the impact of policies and culture on population genetics.
   b. After The Pill and Roe v. Wade, is there genetic selection in the U.S. against traits associated with reproductive choice (whatever those might be)?

8) Feasibility of Personal Genomics as a diplomacy tool (e.g., 23andme "ancestry finder")?

In the next 5 to 10 years, we should see more examples of gene variation linked to social behavioral variation, especially with the kinds of genomic variation that are currently understudied. We will not likely be able to predict with meaningful accuracy an individual's behavior based on an individual's genome or life history. However, at group levels we will have increasing power to assess the probability of given traits and social behaviors. This can have important societal and economic consequences as personal genomics may help optimize matching for mates, jobs, pets, and living environments. As we approach an era of personalized genomics, we will be forced to confront new questions regarding individual rights and responsibilities. Fortunately, as a social species, our large brains have evolved the capacity to navigate changes in the relationship balance between self and others.

Note

Elizabeth A. D. Hammock, Ph.D. is an Instructor of Pediatrics at Vanderbilt University, specializing in genes and

https://doi.org/10.2990/30_2_93 Published online by Cambridge University Press

neural circuits supporting social behavior across the lifespan. She is a graduate of Berry College and received her Ph.D. from Emory University as a scholar with the Center for Behavioral Neuroscience. Previous work has been published in such journals as *Science, Molecular Biology and Evolution,* and *Genes, Brain and Behavior*. Her current research is funded by the National Institutes of Mental Health and a NARSAD Young Investigator Award from The Brain and Behavior Research Foundation.

References

1. G. E. Robinson, C. M. Grozinger, and C. W. Whitfield, "Sociogenomics: Social life in molecular terms," *Nature Reviews Genetics*, 2005, 6(4): 257–270.

2. G. E. Robinson, "Development. Sociogenomics takes flight," *Science*, 2002, 297(5579): 204–205.

3. G. E. Robinson, "Integrative animal behaviour and sociogenomics," *Trends in Ecology and Evolution*, 1999, 14(5): 202–205.

4. J. Clarke et al., "Continuous base identification for single-molecule nanopore DNA sequencing," *Nature Nanotechnology*, 2009, 4(4): 265–270.

5. B. Teague et al., "High-resolution human genome structure by single-molecule analysis," *Proceedings of the National Academy of Sciences*, 2010, 107(24): 10848–10853.

6. T. A. Manolio et al., "Finding the missing heritability of complex diseases," *Nature*, 2009, 461(7265): 747–753.

7. K. Wang et al., "Common genetic variants on 5p14.1 associate with autism spectrum disorders," *Nature*, 2009, 459(7246): 528–533.

8. D. B. Campbell et al., "A genetic variant that disrupts MET transcription is associated with autism," *Proceedings of the National Academy of Sciences*, 2006, 103(45): 16834–16839.

9. T. Walsh et al., "Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia," *Science*, 2008, 320(5875): 539–543.

10. H. Stefansson et al., "Large recurrent microdeletions associated with schizophrenia," *Nature*, 2008, 455(7210): 232–236.

11. J. T. Glessner et al., "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes," *Nature*, 2009, 459(7246): 569–573.

12. M. Bucan et al., "Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes," *PLoS Genetics*, 2009, 5(6): e1000536.

13. M. de Bono et al., "Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli," *Nature*, 2002, 419(6910): 899–903.

14. M. de Bono and C. I. Bargmann, "Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*," *Cell*, 1998, 94(5): 679–689.

15. E. A. Hammock and L. J. Young, "Microsatellite instability generates diversity in brain and sociobehavioral traits," *Science*, 2005, 308(5728): 1630–1634.

16. B. E. Stranger et al., "Relative impact of nucleotide and copy number variation on gene expression phenotypes," Science, 2007, 315(5813): 848–853.

17. L. D. Orozco et al., "Copy number variation influences gene expression and metabolic traits in mice," *Human Molecular Genetics*, 2009, 18(21): 4118–4129.

18. V. Guryev et al., "Distribution and functional impact of DNA copy number variation in the rat," *Nature Genetics*, 2008, 40(5): 538–545.

19. BrainSpan: Atlas of the Developing Human Brain [Internet]. Funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01. ©2011, http://developinghumanbrain.org.

20. H. J. Kang et al., "Spatio-temporal transcriptome of the human brain," *Nature*, 2011, 478(7370): 483–489.

21. M. B. Johnson et al., "Functional and evolutionary insights into human brain development through global transcriptome analysis," *Neuron*, 2009, 62(4): 494–509.

https://doi.org/10.2990/30_2_93 Published online by Cambridge University Press