REVIEW ARTICLE
https://doi.org/10.1038/s41559-017-0411-4

nature ecology & evolution

Genomic tools for behavioural ecologists to understand repeatable individual differences in behaviour

Sarah E. Bengston[1*], Romain A. Dahan[2], Zoe Donaldson[3], Steven M. Phelps[4], Kees van Oers[5], Andrew Sih[6] and Alison M. Bell[7,8]

Behaviour is a key interface between an animal's genome and its environment. Repeatable individual differences in behaviour have been extensively documented in animals, but the molecular underpinnings of behavioural variation among individuals within natural populations remain largely unknown. Here, we offer a critical review of when molecular techniques may yield new insights, and we provide specific guidance on how and whether the latest tools available are appropriate given different resources, system and organismal constraints, and experimental designs. Integrating molecular genetic techniques with other strategies to study the proximal causes of behaviour provides opportunities to expand rapidly into new avenues of exploration. Such endeavours will enable us to better understand how repeatable individual differences in behaviour have evolved, how they are expressed and how they can be maintained within natural populations of animals.

The stunning diversity of behaviour within a species has become a thriving area of research for behavioural ecologists. As a result, we now know that repeatable individual differences in behaviour among animals within populations is ubiquitous. Studies under the umbrella of animal personality (reliable differences in behaviours across contexts or time) and behavioural syndromes (correlated behavioural traits)[1–4] have yielded thousands of publications, particularly during the past decade, and lay a solid foundation for understanding the evolution and effect of repeatable variations in behaviour both theoretically and empirically. Nevertheless, fundamental questions remain unresolved. (1) Why are individuals consistent at all? In other words, why is behaviour not infinitely plastic? (2) Why are some behaviours correlated? And why do correlations sometimes vary among individuals and populations? (3) What explains individual differences in developmental plasticity (effects of earlier experiences on subsequent behavioural tendencies)? Differences in contextual plasticity (effects of current conditions on behaviour)? (4) And finally, why do individuals have different behavioural types? Indeed, knowing how and whether selection acts on consistent among-individual differences in behavioural traits has important implications for our understanding of the maintenance of variation within natural populations, a central problem in evolutionary biology.

Genomics has revolutionized our understanding of evolution, ecology and physiology, yet even with recent advances, the study of animal behaviour has been slower to embrace genomic technologies. One possible reason is that until recently the relevant genetic tools have been out of reach for animal behaviourists fascinated by the behavioural diversity within and among non-model species (Box 1). In addition, repeatable behavioural variation is probably the result of multifaceted, highly dynamic and non-linear epistatic, transcriptional, epigenomic, ontogenetic, neural and metabolic processes[5], which makes it hard to study. Plastic traits such as behaviour present specific challenges for studies at the molecular level: compared with morphological and most life-history traits, behaviour is repeatedly expressed, meaning there can be significant trait plasticity within an individual. Plasticity itself can also vary between individuals[6,7]. The phenotypic gambit[8] and a relative lack of integration across Tinbergen's levels of analysis[9,10] has also slowed progress in this area. Moreover, there is scepticism in some circles about whether we need to study traits (including behavioural traits) at the molecular level at all[11–13], and if the benefits outweigh the considerable costs, both in terms of monetary expense and the training required for proficiency[11,13]. Indeed, some fundamental questions about repeatable individual variation in behaviour do not require expensive forays into the world of genomics. For example, if the researcher is interested in the mechanisms underlying a behaviour, there may be few compelling reasons to incorporate genomics if there are already candidate genes related to the behaviour of interest[14]. In addition, we can learn a lot about behavioural evolution (that is, does behavioural variation reflect genetic or environmental causes?) using standard quantitative genetic 'gene free' approaches (such as cross-fostering or common garden experiments) without incorporating genomics.

However, fundamental questions about behavioural variation can be difficult to resolve without some understanding of genetic or physiological variation, and there are compelling reasons to investigate the mechanisms. For example, both environmental and genetic variation shape behaviour, but whether these effects share overlapping molecular mechanisms remains unclear. As such,

[1]Department of BioSciences, Rice University, Houston, TX, USA. [2]School of Life Sciences, Arizona State University, Tempe, USA. [3]Department of Molecular, Cellular, and Developmental Biology, and Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA. [4]Department of Integrative Biology, University of Texas Austin, Austin, TX, USA. [5]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. [6]Department of Environmental Science and Policy, University of California Davis, Davis, CA, USA. [7]Department of Animal Biology, The University of Illinois, Urbana-Champaign, Urbana, IL, USA. [8]Carl R. Woese Institute for Genomic Biology, The University of Illinois, Urbana-Champaign, Urbana, IL, USA. *e-mail: sbengsto@gmail.com

Box 1 | Perks and perils of the model organism

Although relatively loosely defined[107], model organisms are generally systems with traits predisposing them to be tractable for experimental manipulation. Such traits may include readily reproducing in laboratory settings with relatively short generation times and large clutch or litter sizes, robust embryos that can survive manipulation and share important genes across multiple taxa (for example, with humans). To geneticists and many evolutionary biologists, these models may be little more than functional bags of chromosomes. Yet many model species have highly dynamic and complex behaviours that are of interest to behavioural ecologists focused on repeatable behavioural variation. For example, in *Drosophila melanogaster*, the social context can affect both the aggressive behaviour and fitness of males[108], zebrafish (*Danio rerio*) have innate variation in their response to threats[109], and honey bee workers show some of the most predictable and well-understood behavioural plasticity known[19,110]. Additionally, some model organisms have been useful in discovering well-conserved genetic 'toolkits' for complex traits such as social behaviour[18]. Of course, care must be taken to account for potential behavioural changes associated with many generations of adaptation to laboratory conditions. It may be challenging to interpret adaptive significance of model organism behaviour compared with 'wild' systems. However, researchers interested in causal mechanisms should consider focusing their attention on such promising organisms. Robust techniques and methodologies have already been developed and genomes are well annotated with easily searchable gene ontology databases. Through decades of study in neuroscience and neuroendocrinology, candidate systems and genes of interest for behaviours and behavioural (and synaptic) plasticity are already well described. For a behavioural variation researcher wanting to venture into genomics, perhaps to test a new hypothesis, model organisms may be the easiest point of entry. However, as in any field there are costs to this approach, such as limiting the diversity of taxa seen in the field, and model organisms do not necessarily have the traits of interest to many behavioural ecologists. It may also not be the right direction for students who want to develop their own system to use over the course of their career. Yet researchers should not lightly overlook the value available in a model system.

identifying the genes that contribute to environmentally and genetically mediated variation has major implications for understanding the evolution of phenotypic plasticity[15]. In addition, once we have identified these genes, we can begin to ask fundamental questions about pleiotropy[16], that is, whether the same genes influence multiple traits, and whether these genes are under selection[17]. As we examine the genetic basis of the same 'trait' across species, we can begin to discover whether the molecular mechanisms underlying behavioural variation are deeply conserved in evolution[18]. In studies already using these tools, we can see their value for addressing fundamental questions. For example, in an early microarray paper, the authors were able to track the enormous transcriptomic plasticity (39% of the genes expressed in the brain) that contributes to age polyethism in honey bee workers[19]. It is only a question of time until these tools are applied to address fundamental questions about behavioural variation among individuals within populations. For example, understanding why traits such as aggression and exploratory behaviour are often correlated has been hypothesized to be either the result of pleiotropic interactions or linkage with 'aggressiveness genes' favoured in environments where conspecific aggression is beneficial[2,20]. The underlying causes of such so-called spillovers have rarely been tested, although with knowledge of underlying mechanisms it is in principle an empirically tractable question. Similarly, a fundamental question often asked about repeatable behavioural variation — why is an individual's behaviour consistent rather than infinitely plastic? — is often attributed to the costs of plasticity (but see refs [21,22]). However, measuring the fitness costs of behavioural plasticity has proved difficult[23]. In this Review, we focus on the application of genomic tools to advancing such persistent, fundamental questions. We briefly overview the most common tools as well as their strengths and associated caveats. Additionally, we propose specific hypotheses that can be addressed by their integration, and discuss analytical strategies and the unique role of neurobiology.

Introduction to the tools

For behavioural ecologists venturing for the first time into the unfamiliar territory of using molecular approaches, the process can be daunting. Selecting the right high-throughput genomic technology for a project requires careful consideration of the questions posed, resources available, and of the limitations presented by a given tool and a given system. Contemporary approaches examine three levels of genomic variation: genetic, epigenetic and transcriptional. Genetic variation corresponds to sequence differences. Epigenetic variation is evidenced in differences in the molecular marks on DNA and DNA chaperone proteins, which affect DNA accessibility to transcription factors. Transcriptional variation refers to differences in the amount of RNA generated from a particular DNA locus. Variations at any of these levels are probably not independent from each other. For instance, transcriptional variation is probably tied to genetic or epigenetic variation, making parallel approaches that examine multiple levels of variation a potentially powerful approach.

There are three main categories of molecular tool: (1) genotyping molecular markers, which involve the widespread sequencing of genomic DNA across phenotypes, populations and/or species; (2) comparative gene regulation studies, which characterize transcription and its epigenetic regulation; and (3) genetic manipulations, which directly test function. While not specifically a genomic tool, we will additionally discuss whole genome sequencing, as the presence of a sequenced, assembled and annotated genome can significantly affect how useful specific tools may be, but is not necessarily a trivial task to accomplish. Table 1 presents some practical considerations for using these tools. These are, of course, not the only considerations, and further information relevant to tool selection is considered below.

Genotyping molecular markers for gene association studies. Contemporary genotyping methods involve the widespread genotyping of markers across the genome, and can be used to directly compare genome structure (for example, supergenes) and content of different phenotypes at an unprecedented resolution[24]. These methods involve the targeted enrichment of specific, known sequences (whole exome sequencing, targeted enrichment), or the targeting of randomly distributed restriction sites across the genome (restriction-site-associated DNA sequencing (RAD-Seq); genotyping-by-sequencing (GBS)).

Genome-wide association studies (GWAS) involve searching the genome for marker polymorphisms that are associated with variation in the phenotype of interest. Once markers have been identified, targeted sequencing and/or comparison to a reference genome might reveal the location and identity of genes that lie within loci that segregate with the trait. An advantage of GWAS is that it can be carried out in wild, unmanipulated natural populations, and does not require performing specific crosses. However, considering the great number of comparisons and correlations tested using markers across the genome (for linkage maps, GWAS and so on), the detection of genomic regions in significant associations with trait variation often requires an extremely large sample

Table 1 | Logistic constraints associated with popular high-throughput and experimental techniques

	Requires annotated reference genome	Tissue specific	Developmental stage specific	Cost per sample (US$)[a]	Requires prior knowledge of target sequence?	Must kill organism?	Pooling or keeping samples separate required?	Estimate of independent samples needed for analyses
Sequenced and assembled genome	N/A	No	No	1,500+	No	No	Pooling samples necessary	N/A
Genotyping molecular markers								
RAD-Seq	No	No	No	500+	No[c]	No	May require pooling if starting tissue is small	10s–1,000s+[d]
GBS	No	No	No	500+	No[c]	No	Separate	10s–1,000s+[d]
Exome capture	Yes	No	No	350+	Yes	No	Separate	10s–1,000s+[d]
Comparative gene regulation								
RNA-Seq	No, but recommended	Yes	Yes	500+	No	Yes[b]	May require pooling if starting tissue is small, but usually inadvisable	10+
Methylome	No	Yes	Yes	1,200+	No	Yes[b]	Separate	3–10+
ChIP-Seq	Yes	Yes	Yes	415+	No	Yes[b]	Separate	3–10+
Proteome	No, but recommended	Yes	Yes	3,500+	No	Yes[b]	Separate	3–10+
Genetic manipulations								
CRISPR-Cas9-germline	Yes	No	Must be germline	4,500+[e]	Yes	N/A	N/A	N/A
CRISPR-Cas9 and dCas9-postmitotic	Yes	Yes	No	4,500+[e]	Yes	N/A	N/A	N/A
RNAi	Yes	Sometimes	N/A	50+	Yes	N/A	N/A	N/A

CRISPR, clustered regularly interspaced short palindromic repeats; dCas9, deactivated Cas9; N/A, not applicable. [a]Price per samples estimate for a whole project, including analysis, based on storefront costs of the DNA Sequencing Center at Bringham Young University (https://www.scienceexchange.com/labs/dna-sequencing-center-byu), with the exceptions of proteome (from MS Bioworks; http://www.msbioworks.com/services/protein-works/protein-profiling), CRISPR-Cas9 preps (Transgenic & Gene Targeting Core, University of Utah; https://www.scienceexchange.com/labs/transgenic-gene-targeting-core-utah) and RNAi (ThermoFisher Scientific; http://www.thermofisher.com/us/en/home/life-science/rnai/vector-based-rnai/pol-ii-mir-rnai-vectors.html). [b]These methods require the isolation of specific tissue at specific life stages. For behavioural studies, isolation of the brains will usually be preferable. [c]Prior knowledge of genome allows estimation of density of restriction sites present throughout the target genome, and thus allows a more informed selection of restriction endonuclease based on volume of restriction sites in the target genome. [d]Sample sizes depend on intended use. Population genomics and single-nucleotide polymorphism screening will require a few samples per group, while association studies (GWAS, QTL mapping) require massive sample sizes to attain good statistical power. In some cases, association studied may be done with <100 individuals given stringent conditions (see main text). [e]Price per preparation for transgenic model organism preparation.

size (10,000–100,000+)[24]. A large sample size is also important for accurately estimating allele frequency in the population[25].

Quantitative trait loci (QTL) mapping involves generating crosses between behaviourally divergent individuals and then tracking the segregation of markers linked to phenotypic variation across several generations[26]. The power of QTL mapping is limited by the number of meioses that shuffle associations between markers and the causal alleles, and by the number of loci contributing to variation in the trait and their respective effect sizes; sample sizes of the order of several hundreds to thousands of second-filial-generation individuals are required to detect loci of medium effect[27]. QTL mapping also requires hybrids to be viable and animals to be reared under laboratory conditions for several generations.

Gene association studies via either QTL mapping or GWAS are most likely to be successful when variation is discrete, highly heritable and affected by relatively few loci of large effect. For example, in ruffs there are three alternative male mating morphs under negative frequency-dependent selection. This inversion of a supergene was mapped using only 41 individuals[28]. Unfortunately, most repeatable behavioural variation within populations is continuous, rather than discrete, and is likely to be underlain by thousands of genes of small effect. Therefore, crosses between species or populations with discrete variation might be more tractable for mapping. For example, QTL associated with variation in parenting behaviour between two sister species of mice was found using this approach[29].

Comparative gene regulation. Transcriptional profiling (for example, RNA-Seq) can be used to compare gene expression across phenotypes or conditions in order to identify associations between the expression of specific genes or gene pathways and behavioural phenotypes[30,31]. Transcriptomics is a particularly appealing genomic tool for plastic behavioural traits because measuring gene expression (possibly in combination with quantifying static DNA sequence variation) can reveal how the genome dynamically responds to the environment, including the social environment[32]. For example, studies have compared gene expression between animals that either have or have not experienced a change in their environment that causes changes in behaviour. Such studies have shown that the genome is remarkably dynamic: ~10% of the genome responds to a mating opportunity[33–38], predation risk[39–42], or territorial challenge[43–45].

An advantage of transcriptional profiling over GWAS is that lists of differentially expressed genes can be generated with much smaller sample sizes (of the order of a typical behaviour experiment), and therefore can be an accessible point of entry for behavioural ecologists into the world of genomics. However, one challenge of using transcriptional profiling is that it results in an unbiased list of often

hundreds of differentially expressed genes, requiring thoughtful consideration of how best to move forward with investigation of candidate genes. Lists of differentially expressed genes are a far cry from knowing the specific causal variants underlying phenotypic variation, and results are highly contingent on where tissue is sampled because gene expression is highly cell type-specific. Genes involved in development (that is, with organizational effects) are more likely to be expressed at specific embryonic, larval, or pupal stages (though genes originally described as organizational may also have activational functions in adults[18,46]), and gene expression involved in behavioural traits may vary within the specific neuronal circuits that subserve that behaviour. Given these considerations, it is perhaps not surprising that gene expression is also known to vary between field- and lab-reared organisms[47], which calls for thorough validation of lab-based results in natural populations.

The destructive nature of most tissue sampling (especially for brain) poses challenges for obtaining repeated samples, and might not be realistic for long-term studies of marked individuals, or studies with threatened species. An emerging alternative involves using peripheral proxy tissues such as blood[48,49]. This can be useful, but comes with its own limitations. For example, blood measures of gene expression are generally not relevant to expression within circuits of the nervous system; gene expression varies tremendously even between neuronal cell types and brain regions, so looking at tissue as different as blood is unlikely to reveal transcriptional variation driving behaviour in the brain[50] (though, this may depend on the type of gene[51]). However, blood gene expression is influenced by many of the same factors that regulate brain gene expression, including rearing environment, stress and diet. Viewed as markers for repeatable behavioural variation, rather than as causal contributors to behavioural variation, such studies can be informative.

There is also growing appreciation that gene expression patterns are highly contingent on when tissue is sampled because gene expression can change quickly — on the scale of minutes to hours — and we know very little about the arc of this time course (see C. C. Rittschof and K. A. Hughes, manuscript in preparation). Therefore differences between phenotypes or experimental treatments may exhibit distinct patterns 30 or 120 minutes[52] after a behavioural interaction, for example. The molecular responses to behaviourally relevant stimuli probably involve waves of transcription associated with various types of behavioural plasticity (detecting the stimulus, assessing the stimulus, responding to the stimulus, maintaining a response to the stimulus, recovering from the stimulus and preparing to modify future behaviour after the interaction[53]). Therefore the particular gene expression profile at a specific point in time is just a snapshot of a very dynamic process[54]. Arguably, gene lists produced by a cross-sectional transcriptomic experiment can be difficult to interpret without additional controls to tease apart gene expression associated with movement per se, responding to novelty, responding to any conspecific, versus responding to a same-sex conspecific and so on[55]. Another approach that is likely to be insightful is to measure the time course of gene expression following a stimulus in order to identify the waves of transcription associated with different components of behavioural plasticity[54].

Increasingly, researchers are coupling comparative gene expression to the examination of the mechanisms that regulate chromatin and other epigenetic modifications that influence which genes are expressed and which remain silent. These analyses often examine the methylation state of DNA (bisulfite-converted restriction-site-associated DNA sequencing (BS-RAD-Seq))[56], methylated DNA immunoprecipitation (meDIP), bisulfite sequencing (BS-Seq)[57,58] and pyroseq[59]), the presence of histone modifications and/or transcription factors (chromatin immunoprecipitation (ChIP-Seq)[60]), or chromatin accessibility (assay for transposase-accessible chromatin using sequencing (ATAC-Seq)[61]).

Genetic manipulations. For some research aims, the goal is to find genes to understand how they function to affect the biological process or trait of interest. Given the correlative nature of the aforementioned methods, researchers are increasingly interested in validating their results by manipulating the expression of the gene of interest in order to confirm that it has a causal effect on the trait. Gene expression can be directly manipulated through knock-downs that reduce function, knock-outs that eliminate function, or knock-ins that replace one sequence with another or insert a novel sequence. The phenotypes of the altered organism can be measured, thereby providing a direct test of functionality of targeted candidate sequences. A classical technique with widespread use has been RNA interference (RNAi)[62]. RNAi can modify gene expression, allowing for the targeted testing of pleiotropic effects, providing a clearer view of how specific genes affect the strength of the correlation. RNAi can also be implemented at any life stage, but need not function at every life stage. Applying RNAi in non-model species, particularly in the context of manipulating brain gene expression, is not a trivial effort as it can be unreliable and inefficient until protocols can be modified for new systems. More recently, the clustered regularly interspaced short palindromic repeats-CRISPER associated protein 9 (CRISPR-Cas9) system has been developed for genome editing[63]. It is extremely efficient for gene knock-outs, and is likely to be a very accessible tool for species with amenable reproductive biology, such as resilient embryos that can be manipulated. For systems where this is not possible, post-mitotic CRISPR-Cas systems are a potential option for localized gene manipulation[64,65]. It also allows the insertion of novel sequences into the target genome, providing a versatile tool for functional genomics; however, these knock-ins are orders of magnitude less efficient than targeted deletions, and are probably a poor choice for an initial causal study in a non-model system. There are, of course, limitations such as difficulties in confirming the correct target sites were reached and mediating off-site effects (as reviewed in ref. [66]). Genome editing also requires one to be able to rear the organism in the lab and, preferably, to affect germ-line cells, so that the edited genome may be passed through to subsequent generations (discussed further in 'The role of neuroscience', below).

Whole genome sequencing. Understanding the genomic basis of behavioural differences often involve sequencing the entirety of a species' genome. Sequencing a reference genome for a study system will greatly increase the power and ease of genomic analyses described above. Reference genomes enable researchers to ask more advanced questions about the genomic and epigenomic variation underlying behavioural differences, such as identifying regulatory regions, both *cis*- and *trans*-, involved upstream of differentially expressed genes identified through RNA-Seq[67]. In addition, whole genome sequencing (WGS) can be used to identify patterns of selection and/or divergence between closely related species, populations, or behavioural types, such as relaxed selection in a large non-recombining region associated with social structure in fire ants[68], or genome differentiation between diverging behavioural types of malaria mosquitoes[69,70].

WGS does not hold the answers to all genomics questions, and as with all the tools described here, requires careful consideration. Obtaining even a single, high-quality reference genome can be an arduous task that is time- and resources-intensive. Genome sequencing, and indeed most next-generation sequencing, relies on short reads, usually followed by assembly steps to yield longer sequences. One of the best measures of sequence quality is coverage (or depth), that is, how many sequence reads include a particular locus. To obtain high coverage, and therefore a high-quality sequence, the amount of starting material is important, including the size of the target genome and the number of starting samples. Coverage will also usually be lower at polymorphic sites, where

high levels of heterozygocity occur, as well as in highly repetitive regions, where assembly can be unreliable. Newer sequencing methods involve longer reads of single molecules, which may help resolve poor assembly of highly repetitive regions. However, these methods can be error-prone and require the complementary use of short-read sequencing for error corrections. WGS projects, as well as genome-wide projects involving assembly (such as RAD-Seq and RNA-Seq) often require tradeoffs between coverage and sample size (for example, the decision to pool samples to obtain higher quality sequences, as in ref. [69]), where it may be beneficial to have lower-quality thresholds (for example reducing the target coverage from 20× to 10×), to increase the available sample sizes.

With a known reference genome and identified candidate genetic regions, targeted sequencing approaches can be used to detect genetic factors involved in repeated inter-individual differences with more precision. These techniques can take advantage of polymerase chain reaction (PCR) technology to increase sequence quality without sacrificing sample size. For quantitative gene expression projects, the non-linear amplification introduced by many cycles of PCR can distort results, a problem that can be avoided by using simple techniques such as emulsion PCR or quantitative PCR.

Aligning tools with questions

For many questions about repeatable behavioural variation, the hope is that using molecular methods will generate data that provide novel inferences about how and why behavioural variation exists. This may come in the form of inductive inferences from patterns — are the same genes associated with aggressiveness and exploration? — or in trying to connect genes directly to neurobiological, endocrine or physiological processes that affect behavioural responses. In all of these cases, a priori understanding of which genomic or molecular data are sufficient to test hypotheses is critical because the same genomic technique may not advance every question. Each tool has a unique set of strengths, weaknesses and applications to questions of interest. For example, questions about within-individual change such as developmental and contextual plasticity may be best approached using tools that measure genome-wide expression (RNA-Seq). In model systems, these questions could also be approached through genome and/or gene expression editing techniques to manipulate the expression of genes of interest. If one or several genes are suspected to regulate the correlation of many behaviours, for example, a knock-out experiment may be appealing[71]. Questions more focused on between-individual variation may be better served through massive parallel sequencing techniques that can compare the content or structure of the genome. For example, exome capture may be useful in identifying single-nucleotide polymorphisms associated with different behavioural types in a GWAS study, while RAD-Seq may be more useful in comparing the genomic structure between behavioural types or between species via QTL mapping. Table 2 gives example hypotheses highlighting how the different tools could be applied to five fundamental questions about repeatable behavioural variation, behavioural plasticity and trait co-variances, thereby providing a framework to select the right tool for the question at hand. This table is not intended to be a complete list of hypotheses or each tool's potential, but rather a starting point for those familiarizing themselves with new tools. Below, we highlight five examples to illustrate the application of these tools.

Using GWAS to understand why there are behavioural types. A proximal hypothesis for why there is behavioural variation among individuals within natural populations is that different behavioural types of individuals have different variants of a gene that influences behaviour. This question was addressed in a study of the silver alpine ant *Formica selysi*, which applied GWAS between social morphs to identify a large Mendelian supergene associated with variation in social structure in colonies. This supergene is predictive of the non-sibling queen tolerance of the colony (a key aspect of the colony's personality[72]), thus influencing if a colony is monogynous or polygynous[73]. It is plausible that once the genes relating to behavioural variation have been found, genome data can be used to address whether balancing selection is maintaining the genetic variation, thereby providing an ultimate answer to the same question[74].

Applying RNA-Seq to understand why individuals behave consistently. One possible reason why individuals might maintain a behavioural type is because it's too costly to switch to a different one. According to this line of reasoning, the shift between behavioural types requires dramatic transcriptional change. This is expected to be costly, as it results not only in transcriptional change but also downstream reorganization of signalling cascades and physiological properties, requiring both time and energy. A simple prediction, then, is that there are fewer transcriptional changes associated with behaviours that are highly flexible within individuals, but large transcriptional changes associated with the shift between more stable behavioural types. Consistent with this idea, and based on emerging evidence that it is energetically costly to change gene expression significantly, sometimes with fitness consequences[75], the relatively stable, permanent change in behaviour of honeybee workers from nursing to foraging is associated with a 39% difference in gene expression. However, changes in gene expression are much lower when workers switch between less stable and more flexible occupations, such as undertaker and guard[19,76]. An alternative view is that maintaining a behavioural type is similar to maintaining homeostasis. In that case, nonplastic individuals, that is, those whose behaviour does not change dramatically in response to the environment, might experience large fluctuations in gene expression that are related to maintaining their behavioural type (Table 2). Experiments that measure gene expression on a genome-wide scale can distinguish between these two hypotheses.

Applying RNA-Seq to understand why individuals differ in behaviour. The gene expression profile of individuals with different behavioural types, such as between genetic lines selected for high or low levels of a particular behaviour, or between alternative behavioural phenotypes, can be compared. Differences in expression between behavioural types could be caused by genetic variation, epigenetic modifications or developmental plasticity. Arguably, this experimental design is better suited to answering questions about the molecular causes and correlates of individual variation, rather than questions pertaining to changes in gene expression in response to given stimuli, because detected differences in gene expression could reflect molecular processes involved in maintaining a particular neural structure/function (that is, maintaining plasticity), rather than generating a neurogenomic state[77,78]. This means gene expression changes may reflect suites of traits associated with behavioural variation that are also often of interest to behavioural ecologists (for instance, pace-of-life syndrome)[79]. For example, genes that are differentially expressed between alternative phenotypes that differ in aggressiveness (for example, sneaker/satellite males verses territory holders) probably reflect processes involved in maintaining the molecular machinery associated with morphological and life-history differences between the phenotypes, such as reproductive maturation[53].

Another tactic is to present individuals with a behaviour-relevant stimulus, record their behaviour, measure gene expression in response to the stimulus and include individual behaviour in the analysis of gene expression data. For example, in sticklebacks, the expression of differentially expressed genes in response to an intruder was correlated at the individual level with levels of aggressiveness, suggesting that differences in gene expression reflect, in part, individual differences in behaviour[44]. Identifying genes changing in expression, that is, as an upstream promoter or

downstream target, will perhaps better advance our understanding of the architecture of behavioural types, as discussed below. Another possibility is to compare different behavioural types in a response experiment with the same control and to ask if there are more differentially expressed genes between one extreme behavioural type and the unexposed control of some type versus another behavioural type and its control.

Applying RNA-Seq to understand why traits are correlated. When individual differences in behaviour are correlated in response to different ecologically relevant stimuli (for example, a potential mate, competitor, predator, new environment and so on), genes that are differentially expressed in response to both stimuli are plausible molecular causes or correlates of the behavioural syndrome. Correlated behaviours can be expressed through modulation of the expression of the same causative loci in different contexts. Comparative gene expression analyses involving individuals responding to different cues can reveal the role of up- and down-regulation of the same genes in different contexts. For example, the expression of genes in the brain in stickleback fish changes both in response to courtship and aggression; albeit in different directions. This suggests that such genes are involved in both aggressive and courtship behaviours, causing behavioural traits to be correlated[80].

Using gene association studies to understand why traits are correlated. One possible explanation for behavioural syndromes is that different behaviours are influenced by the same genes, or set of physically linked loci. If this is the case, then the different behavioural traits will map to the same genomic region(s) in gene association studies. For example, the close proximity of two QTLs participating in male display behaviour and female preference in *Lapaula* crickets explains the correlation of both of these traits[26]. At the individual level, correlated traits of white-throated sparrows involved in parental care, plumage and aggression were linked to co-expression of genes located in a chromosomal inversion[81].

Of course, no one tool is the magic bullet to resolve the persistent questions in repeatable behavioural variation research. Instead, the most important insights, as well as the most successful research avenues, are likely to come when tools with complementary strengths are integrated (Box 2). For example, by integrating RNA-Seq and ChIP-Seq, one study demonstrated rapid and dramatic epigenomic plasticity in response to social interactions in three-spined sticklebacks. This was facilitated by integrating brain gene expression data with a transcriptional regulatory network, and linking gene expression to changes in chromatin accessibility (Fig. 1)[54].

Inferring the function of genes

Twenty years ago, sequencing the human genome was one of the most ambitious scientific endeavours ever attempted. However, genomic technology has advanced quickly and has rapidly outpaced our computational capabilities, creating challenges in interpreting genomic data. Yet, there are strategies that work broadly and are applicable to research in repeatable behavioural variation and behaviour generally, such as functional annotation and gene classification. This is only a subsampling of the analytical tools (most of which are covered in a quality genomic sciences primer; for example, ref. [82]) and considerations for inferring gene function (see also ref. [46] for a more thorough discussion of these topics).

A primary goal of any genome sequencing project is to classify genes into putative functional families. This allows for necessary comparisons, perhaps to look for genes overrepresented or underrepresented compared with other genomes. While alignment (where a reference genome is available) or assembly (where one is not) may be a first step in functional annotation, this method is insufficient and error prone, meaning further steps are required[83]. Using software such as BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), amino acid similarity to previously annotated genes can be compared. However, it is not uncommon for one-third to one-half of the genome to remain unclassified after such an analysis. Numerous databases that are often discipline-specific have been established to classify these protein domains. Currently, however, no such database exists for behavioural genomic data.

Annotation based on molecular function is insufficient to describe or predict biological function. An annotation cannot take neo-functionalization directly into account, where the physiological

Box 2 | Integrating tools

As tools increase in popularity and accessibility, it will become increasingly possible to use them in complementary ways. This is attractive because each tool has strengths and weaknesses, and investigations of the toughest questions about repeatable individual variation at the molecular level are likely to be best served by the application of multiple genomic tools in creative and complementary ways. Here, we offer examples of how studies of behavioural variation are likely to benefit from the simultaneous application of multiple tools.

Integrating RNA-Seq and GWAS/QTL to find genes related to behavioural variation. One of the main drawbacks to gene association studies is the large sample size needed to narrow the window harbouring key genetic variants. An obvious way to reduce the search space and hence improve power to detect loci is to prioritize regions of the genome that harbour genes that are differentially expressed, based on results from RNA-Seq studies.

Integrating RNA-Seq and epigenomics (ChIP-Seq, ATAC-Seq, methylation profiling) to understand behavioural plasticity. Many of the outstanding questions about repeatable behavioural variation are to do with behavioural plasticity. Transcriptomic profiling is especially well suited for investigating behavioural plasticity at the molecular level, but our questions are increasingly focused on upstream regulators of transcriptional plasticity, especially insofar as they might tell us about the causes of variation in plasticity. If we can identify key regulatory elements that govern changes in gene expression (for example, histone modifications, transcription factor binding sites, methylation and chromatin accessibility), then we can start asking questions about genetic variation in those elements, which might be related to individual differences in plasticity.

Integrating gene association studies with epigenomics to explore constraints on plasticity. Individuals may be behaviourally constrained if their behavioural type is genetically or epigenetically influenced. Genetic markers generated via GBS or RAD-Seq can be used to perform a GWAS or to identify QTLs that may be associated with a specific behavioural type, identifying genetic regions in linkage disequilibrium with causative alleles involved in constraining plasticity, as discussed above. On the other hand, epigenetic modifications, such as histone modifications or methylation differences, can be identified using ChIP-Seq, reduced representation bisulfite sequencing (RRBS), whole genome bisulfite sequencing (WGBS) or pyrosequencing, and have been shown to affect behaviour[58]. Functional tests can be used to confirm results. For example, histone modifications in carpenter ants have been shown to directly affect foraging behaviour of workers. If these modifications are altered, the behaviour of workers is also altered, suggesting a causal link between chromatin state and constraints on behavioural plasticity[111].

Table 2 | Applying tools to questions

(a)	Sequenced and assembled genome	Genotyping molecular markers			Comparative gene regulation			
		RAD-Seq/ GBS	Candidate gene region	Exome capture	RNA-Seq	Methylome	ChIP-Seq	Proteome
Why is there consistency?	Consistent interindividual differences in behaviour result from major detectable changes in the genomes	Consistent interindividual differences in behaviours result directly from interindividual sequence and/ or structure differences in association with genomic markers	Consistent interindividual differences in behaviours result directly from interindividual sequence differences in a given candidate genetic region	Consistent interindividual differences in behaviours result directly from interindividual differences in protein sequences	Consistent behaviours are the result of conserved genetic networks being expressed in the same circumstances	Comparing the methylome of individuals among behavioural types can explain the degree of consistency within/ between behavioural types	Consistent behaviours result from epigenetic constraints preventing the expression of alternative phenotypes	See RNA-Seq
Why are traits correlated?	X	Correlated behavioural traits are expressed by pleiotropic or closely linked genes	See RAD-Seq/ GBS, testable with more precision. This requires the prior knowledge of candidate regions	Correlated traits are expressed via pleiotropic changes	Correlated traits are governed by genes involved in the same transcription regulatory network, enhanced by the same TFs	Correlated traits are governed by genes correlated in their methylation states	Genes coding for correlated traits become accessible through epigenetic changes involving the same TFs and/ or histones	Expression of correlated traits involve multiple protein variants folded from the same amino acid chains through post-translational modifications
Why are there differences in developmental plasticity?	X	Differences in the level of plasticity for a trait result from differences in the sequence and/or structure of the genome	Differences in the level of plasticity in behaviour result directly from sequence differences in a given candidate genetic region	Differences in the level of plasticity in behaviours result directly from differences in protein sequences of each morph	Differences in the level of plasticity in behaviour result from differences visible through transcriptomes (e.g. changes in the transcript regulatory networks, changes in TF binding sites)	Differences in the level of plasticity in behaviour scale with differences in the variance in levels of DNA methylation between plastic and nonplastic morphs following critical developmental period	Differences in the level of plasticity in behaviour result from epigenetic differences at the critical developmental time, rendering genomic regions involved in plasticity inaccessible	Differences in the level of plasticity in behaviour result from differences in maternal effect proteins available at critical times (may also be visible through transcripts)
Why are there differences in contextual plasticity?	X	Differences in the level of plasticity for a trait result from differences in the sequence and/or structure of the genome	Differences in the level of plasticity in behaviour result directly from sequence differences in a given candidate genetic region	Differences in the level of plasticity in behaviours result directly from differences in protein sequences of each morph	Differences in the level of plasticity in behaviour result from differences visible through transcriptomes (e.g. changes in the transcript regulatory networks, changes in TF binding sites)	Differences in the level of plasticity in behaviour scales with differences in the level of variance in DNA methylation following a cue	Differences in the level of plasticity in behaviour result from differences in the scale and shape of epigenetic modifications between morphs	See RNA-Seq

Continued

Table 2 | Applying tools to questions (Continued)

(a)	Sequenced and assembled genome	Genotyping molecular markers			Comparative gene regulation			
		RAD-Seq/ GBS	Candidate gene region	Exome capture	RNA-Seq	Methylome	ChIP-Seq	Proteome
Why are there differences in behavioural type?	There are major genomic differences between types	Behavioural types are the result of genetic changes in sequence and/ or structure in the genome	Behavioural types are the result of sequence change in specific candidate genetic regions. Detectable with more power than in genome-wide methods	Behavioural types are the result of sequence change in the protein-coding sequences between morphs	Behavioural types are the result of major changes in transcription regulatory networks between types	Behavioural types are the result of profound and maintained differences in methylation, especially around genes involved in the expression of divergent behaviours	Behavioural types are the result of profound and maintained epigenetic differences between types	Differences in post-translational processes contribute to differences in behavioural types

(b)	Sequenced and assembled genome	Genetic manipulations		
		CRISPR		RNAi
		Germ-line	Postmitotic	
Why is there consistency?	Consistent interindividual differences in behaviour result from major detectable changes in the genomes	Affecting the genome sequence of loci identified as causing consistent behaviours affects the behaviours as well	Modulating the expression of genes involved in the expression of consistent behaviours affects the behaviour	Modulating the expression of genes involved in the expression of consistent behaviours affects the behaviour
Why are traits correlated?	X	Correlated traits are affected by pleiotropic genes. Thus, affecting the function of genes involved changes the expression of all correlated behaviours	Correlated traits are affected by the expression of pleiotropic genes in the different tissues and/ or at different life stages	Correlated traits are affected by the expression of pleiotropic genes in the different tissues and/or at different life stages
Why are there differences in developmental plasticity?	X	Modifying genic regions involved in the plastic response to an environment will significantly alter the plasticity of a trait	Preventing the transcription of genes with modular expression across a reaction norm will render a plastic behavioural trait canalized	Blocking the translation of maternal effect mRNA identified as being involved in differences in plasticity will canalize the expression of a trait
Why are there differences in contextual plasticity?	X	Modifying genic regions involved in the plastic response to an environment will significantly alter the plasticity of a trait	Preventing the transcription of genes with modular expression across a reaction norm will render a plastic behavioural trait canalized	Knocking down the transcription of genes with modular expression across a reaction norm will render a plastic behavioural trait canalized
Why are there differences in behavioural type?	There are major genomic differences between types	Genetic modifications of divergent sequences from its state in one type to its state in another type will result in a reduction of the phenotypic differences between types	Modulating the gene expression of genes with differences in expression between types, especially in the brain, will result in changes in the phenotypes of behavioural types	Modulating the gene expression of genes with differences in expression between types, especially in the brain, will result in changes in the phenotypes of behavioural types

Shown are possible ways to apply tools to five outstanding questions about repeatable behavioural variation. Specific tools may have limitations making them unsuitable for specific taxa or approaches (see Table 1 and main text). While methodologically very different tools, the application of RAD-Seq data is similar to that of GBS data (and of any other genotyping method involving random markers in the genome), therefore only one column was included describing the application of both. TF, transcription factor.

function of a gene has evolved, or where phenome-level traits have split from one gene to several[82]. Yet, there is still highly conserved gene function across most animal taxa, meaning that understanding the ontology of genes still holds value[53] (but see the phenolog concept[84]). Projects such as the Gene Ontology Consortium (GOC; http://www.geneontology.org) work to try and find unifying functions of genes and gene products across eukaryotes[85]. As cell biologists and physiologists have pioneered much of the gene ontology work, the annotations are biased as such. This can make direct inferences about behavioural variation difficult. While smaller than the GOC, one resource of particular value to behaviour researchers may be www.geneweaver.org, which includes the functional effects on behaviour when available and has consolidated much of the published behavioural genomics work[86].

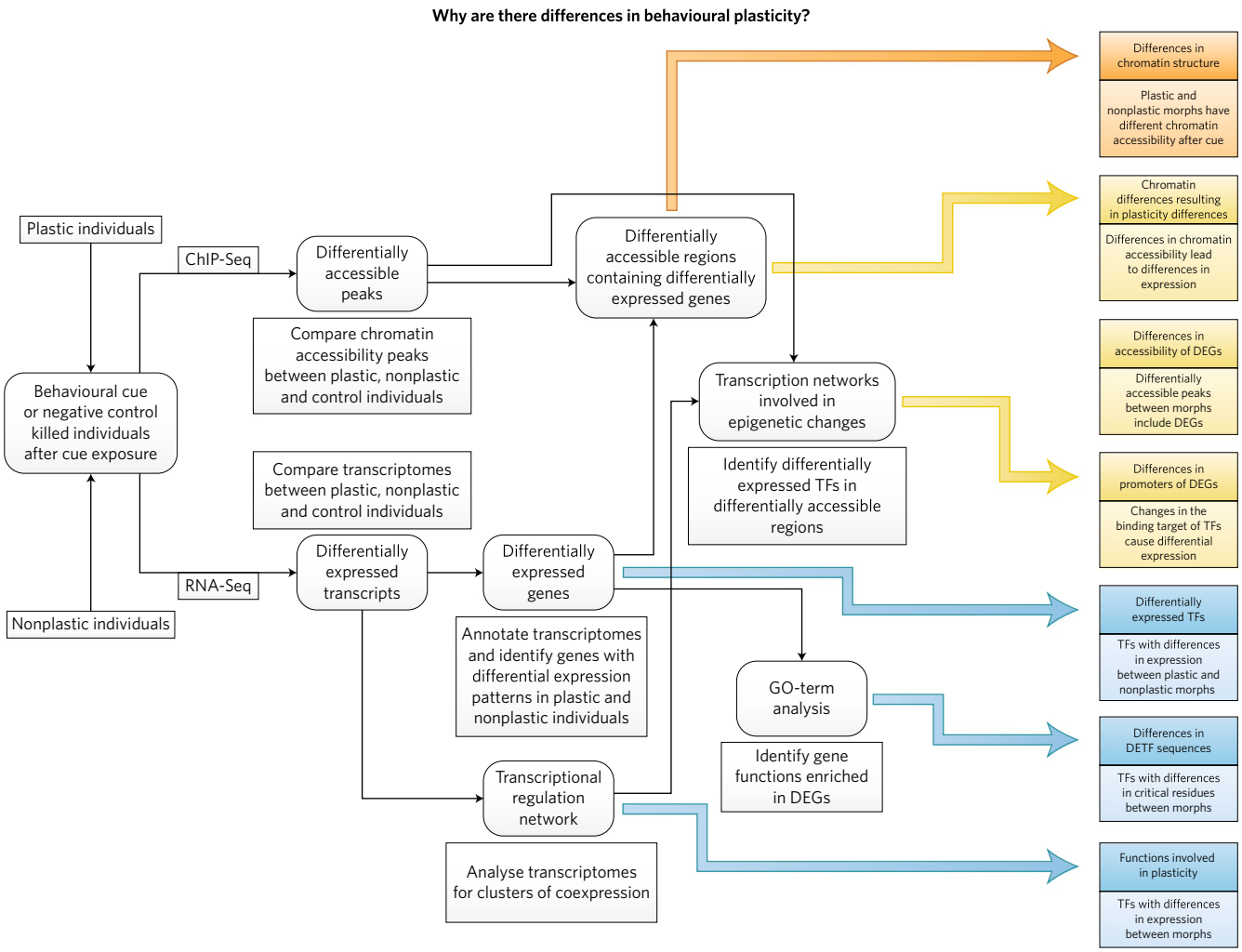

Fig. 1 | A flowchart of a hypothetical study emphasizing the benefit of integrating multiple tools to understand repeatable behavioural variation; in this case, combining RNA-Seq and ChIP-Seq to investigate differences in behavioural plasticity. The orange arrow and boxes represent what can be inferred from a ChIP-Seq protocol alone; blue arrows and boxes represent what can be inferred from a RNA-Seq protocol alone. Golden arrows and boxes highlight what can be inferred by combining both tools to obtain complementary information. DEG, differentially expressed genes; GO, gene ontology; DETF, differentially expressed transcription factor. Inspired by methods used in ref. [54].

The dearth of resources for relating the function of genes to processes of interest to behavioural ecologists is apparent. This results in some of the 'scepticism' about what insights come from behavioural variation studies that yield lists of up- versus down-regulated genes, but no functional significance of these gene expression patterns (although the pattern is sometimes important irrespective of the gene identities[87]) and no clear next step. This problem may be resolved by deploying other strategies or applying other tools to further explore genes of interest. Simply knowing the number of genes changing their expression may not be inherently helpful. It may seem an obvious prediction that big switches between behaviours are a result of big transcriptional changes. Perhaps this may be proposed as a hypothesis for why individuals stay consistent — because big switches are mechanistically difficult. However, this may be an over-simplification, as the number or size of a transcriptional change may not reflect the ultimate 'cost' to the organism.

Creating databases with behaviour-centric protein domains or gene ontologies is an alluring, but a potentially impractical path for pioneering behavioural ecologists. However, understanding repeatable behavioural variation is inherently integrative, with researchers putting forth a significant number of hypotheses for why this variation has emerged and been maintained. Many of these hypotheses holistically incorporate different trait types; for example, it has been suggested that physiology (fast–slow physiology[88,89]), life-history strategy (pace-of-life hypothesis[79]), speed–accuracy trade-offs in cognitive differences[90] and variation in immune function[91] may drive observed behavioural variation. Using a technique such as a functional enrichment analysis could be used to further this integration. For example, if gene ontology terms are found to be related to metabolism, neurotransmission or immunity, this may offer support for particular hypotheses about behavioural variation. A practical approach may be to integrate better with genomicists or other researchers already developing gene ontology databases.

Neurogenetic basis of behavioural diversity

The path from genome to phenome passes through the structure and function of the nervous system, where changes in gene expression influence the development of circuit connectivity or transiently modulate cellular properties. Because the brain is characterized by its exquisite heterogeneity of cell types, a satisfying link between genetic mechanisms and behavioural outcomes requires attending to differences in specific brain regions and circuits. As a result, the

'where' and 'when' of gene expression are critical considerations. While this seems daunting to most behavioural ecologists, there are good suggestions for how best to incorporate neuroscience when considering the biological mechanisms that contribute to repeatable behavioural variation.

One of the main reasons to incorporate neuroscience into studies of behavioural variation is to improve the power of genomic approaches by focusing efforts on relevant circuits and brain regions. Performing an RNA-Seq study with an entire brain could reveal important differences (Table 2), but because the sequencing will include a majority of transcripts from brain regions unrelated to the behaviour of interest, the effects of important genes may be masked, or may require an increase in sequencing effort to detect. There are well-known circuits for most dimensions of behaviour that interest behavioural variation researchers, including aggression, boldness and energy balance, and the past decade has witnessed substantial advances in identifying the homologues of brain regions across vertebrates[92–94]. Knowing these circuits can assist in our understanding of trait correlations, as well. For example, if the same nodes within these circuits are involved in multiple behaviours, correlations between traits might emerge from variation in gene expression across these common neural structures[95,96].

Although the a priori choice of brain regions based on homology is powerful and general, an alternative approach is to use an unbiased method to identify brain regions that differ in function between individuals with alternative phenotypes. One simple but neglected method is to examine the metabolic activity of brain regions by staining for cytochrome oxidase[97], the rate-limiting enzyme in oxidative phosphorylation that changes in response to use that spans days or weeks — a time scale well suited to behavioural variation research (C. C. Rittschof and K. A. Hughes, manuscript in preparation). Brain regions exhibiting differences in metabolism could be useful targets for subsequent studies of the transcriptome or epigenome.

Another, more common, approach is to examine the expression of specific genes that are expressed in response to recent neural activity, known as 'immediate early genes' (IEGs). By evoking a specific behaviour from an animal, relevant brain regions become activated and, with modest delay, express IEGs. However, commercial antibodies for detecting IEG products tend to transfer poorly to non-mammalian species (though this complication can be circumvented by using mRNA in situ hybridization[98]). Thus, a promising new approach is to examine the transient phosphorylation of actively translating ribosomes[99], marks that are enriched by neural activity and highly conserved across taxa. Sequencing the mRNA being actively translated by these phosphorylated ribosomes also allows the researcher to identify neurons of interest, for instance via their expression of specific neurotransmitter-related genes. Although still preliminary in its application, this tool may be useful for researchers studying behavioural variation.

Once specific genes and brain regions have been identified, the next logical step is to manipulate gene expression, such as using CRISPR-Cas9 to create knock-outs and observe their behavioural consequences[100–103]. However, the absence of a gene product throughout the animal's life provides a relatively poor model of naturally occurring genetic or transcriptional variation, and biological compensation can mask the effects of some gene knock-outs. Thus, more relevant approaches may be ones in which levels of gene transcription can be manipulated in a regionally and temporally refined manner within the nervous system. RNAi[104], modified Cas9-fusion proteins (see Table 2)[102,105], and CRISPR-mediated site-specific epigenetic modifications[106] represent tools for achieving localized manipulations of gene function. In the lab, these can be introduced via replication-deficient viral vectors. However, transitioning these technologies to the field may require additional safety considerations.

The decision to incorporate neuroscience and mechanistic studies does require consideration of what will be gained from doing so. In particular, doing so may help reveal whether there are general principles that explain the neural mechanisms underlying repeatable behavioural variation across species. Additionally, focusing on key brain areas for gene expression may provide cleaner data for more targeted hypothesis testing, as well as a better understanding of how genetic variation exerts its effect on behavioural phenotype.

Is behavioural genomics right for you?

Entering the world of behavioural genomics comes with a considerable set of challenges and considerations. For example, organisms with large genomes, that are polyploid and have a large number of repetitive elements are likely to pose challenges for studies that require an assembled genome, and certain organisms are going to be easier to manipulate than others (Box 1). Moreover, it is worth carefully considering whether the benefits of using these tools are likely to outweigh their considerable costs. As argued here, we clearly think the answer is 'yes', and that there are apparent opportunities for research investigating repeatable behavioural variation to benefit from incorporating modern genomic techniques (Box 2). We hope that this is an appealing direction to many, with the possibility of pioneering new analytical methods and taking a leadership position in directing this field forward.

Received: 14 May 2017; Accepted: 10 November 2017;
Published online: 12 February 2018

References

1. Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K. & van Noordwijk, A. J. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Anim. Behav.* **64**, 929–938 (2002).
2. Sih, A., Bell, A. & Johnson, J. C. Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19**, 372–378 (2004).
3. Pruitt, J. N. & Keiser, C. N. The personality types of key catalytic individuals shape colonies' collective behaviour and success. *Anim. Behav.* **93**, 87–95 (2014).
4. Bengston, S. E. & Dornhaus, A. Be meek or be bold? A colony-level behavioural syndrome in ants. *Proc. R. Soc. B* **281**, 20140518 (2014).
5. Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* **15**, 22–33 (2014).
6. Stamps, J. A. & Biro, P. A. Personality and individual differences in plasticity. *Curr. Opin. Behav. Sci.* **12**, 18–23 (2016).
7. Dingemanse, N. J. & Wolf, M. Between-individual differences in behavioural plasticity within populations: causes and consequences. *Anim. Behav.* **85**, 1031–1039 (2013).
8. Grafen, A. in *Behavioural Ecology* 2nd edn (eds Krebs, J. & Davies, N.) 62–84 (Blackwell, Oxford, 1984).
9. Bateson, P. & Laland, K. N. Tinbergen's four questions: an appreciation and an update. *Trends Ecol. Evol.* **28**, 712–718 (2013).
10. Stamps, J. Behavioural processes affecting development: Tinbergen's fourth question comes of age. *Anim. Behav.* **66**, 1–13 (2003).
11. Travisano, M. & Shaw, R. G. Lost in the map. *Evolution* **67**, 305–314 (2013).
12. Paaby, A. B. & Rockman, M. V. The many faces of pleiotropy. *Trends Genet.* **29**, 66–73 (2013).
13. Zuk, M. & Balenger, S. L. Behavioral ecology and genomics: new directions, or just a more detailed map? *Behav. Ecol.* **25**, 1277–1282 (2014).
14. Fitzpatrick, M. J. et al. Candidate genes for behavioural ecology. *Trends Ecol. Evol.* **20**, 96–104 (2005).
15. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. (Oxford University Press: New York, 2003).
16. Saltz, J. B., Hessel, F. C. & Kelly, M. W. Trait correlations in the genomics era. *Trends Ecol. Evol.* **32**, 279–290 (2017).
17. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
18. Rittschof, C. C. et al. Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. *Proc. Natl Acad. Sci. USA* **111**, 17929–17934 (2014).
19. Whitfield, C. W., Cziko, A.-M. & Robinson, G. E. gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296–299 (2003).
20. van Oers, K. & Mueller, J. C. Evolutionary genomics of animal personality. *Phil. Trans. R. Soc. B* **365**, 3991–4000 (2010).

21. Stamps, J. A. & Frankenhuis, W. E. Bayesian models of development. *Trends Ecol. Evol.* **31**, 260–268 (2016).
22. Sih, A. et al. Animal personality and state–behaviour feedbacks: a review and guide for empiricists. *Trends Ecol. Evol.* **30**, 50–60 (2015).
23. Snell-Rood, E. C. An overview of the evolutionary causes and consequences of behavioural plasticity. *Anim. Behav.* **85**, 1004–1011 (2013).
24. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95–108 (2005).
25. Slate, J. From beavis to beak color: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution* **67**, 1251–1262 (2013).
26. Shaw, K. L. & Lesnick, S. C. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl Acad. Sci. USA* **106**, 9737–9742 (2009).
27. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, 1998).
28. Küpper, C. et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
29. Bendesky, A. et al. The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434–439 (2017).
30. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* **32**, 690–703 (2015).
31. Anholt, R. R. H. et al. The genetic architecture of odor-guided behavior in *Drosophila*: epistasis and the transcriptome. *Nat. Genet.* **35**, 180–184 (2003).
32. Bell, A. M. & Robinson, G. E. Behavior and the dynamic genome. *Science* **332**, 1161–1162 (2011).
33. Lawniczak, M. K. & Begun, D. J. A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. *Genome* **47**, 900–910 (2004).
34. Mack, P. D., Kapelnikov, A., Heifetz, Y. & Bender, M. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 10358–10363 (2006).
35. Carney, G. E. A rapid genome-wide response to *Drosophila melanogaster* social interactions. *BMC Genom.* **8**, 288 (2007).
36. Cummings, M. E. et al. Sexual and social stimuli elicit rapid and contrasting genomic responses. *Proc. R. Soc. B* **275**, 393–402 (2008).
37. McGraw, L. A., Clark, A. G. & Wolfner, M. F. Post-mating gene expression profiles of female *Drosophila melanogaste*r in response to time and to four male accessory gland proteins. *Genetics* **179**, 1395–1408 (2008).
38. Fraser, B. A., Janowitz, I., Thairu, M., Travis, J. & Hughes, K. A. Phenotypic and genomic plasticity of alternative male reproductive tactics in sailfin mollies. *Proc. R. Soc. B* **281**, 20132310 (2014).
39. Mori, T. et al. Genetic basis of phenotypic plasticity for predator-induced morphological defenses in anuran tadpole, *Rana pirica*, using cDNA subtraction and microarray analysis. *Biochem. Biophys. Res. Commun.* **330**, 1138–1145 (2005).
40. Sanogo, Y. O., Hankison, S., Band, M., Obregon, A. & Bell, A. M. Brain transcriptomic response of threespine sticklebacks to cues of a predator. *Brain Behav. Evol.* **77**, 270–285 (2011).
41. Becks, L., Ellner, S. P., Jones, L. E. & Hairston, N. G. The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics. *Ecol. Lett.* **15**, 492–501 (2012).
42. Lavergne, S. G., McGowan, P. O., Krebs, C. J. & Boonstra, R. Impact of high predation risk on genome-wide hippocampal gene expression in snowshoe hares. *Oecologia* **176**, 613–624 (2014).
43. Alaux, C. et al. Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc. Natl Acad. Sci. USA* **106**, 15400–15405 (2009).
44. Sanogo, Y. O., Band, M., Blatti, C., Sinha, S. & Bell, A. M. Transcriptional regulation of brain gene expression in response to a territorial intrusion. *Proc. R. Soc. B* **279**, 4929–4938 (2012).
45. Rittschof, C. C. & Robinson, G. E. Manipulation of colony environment modulates honey bee aggression and brain gene expression. *Genes Brain Behav.* **12**, 802–811 (2013).
46. Rittschof, C. C. & Robinson, G. E. in *Current Topics in Developmental Biology*, Vol. 119 (ed Orgogozo, V.) 157–204 (Academic: Cambridge, 2016).
47. Jandt, J. M., Thomson, J. L., Geffre, A. C. & Toth, A. L. Lab rearing environment perturbs social traits: a case study with *Polistes* wasps. *Behav. Ecol.* **26**, 1274–1284 (2015).
48. Tylee, D. S., Kawaguchi, D. M. & Glatt, S. J. On the outside, looking in: a review and evaluation of the comparability of blood and brain '-omes'. *Am. J. Med. Genet. B* **162**, 595–603 (2013).
49. Nikolova, Y. S. & Hariri, A. R. Can we observe epigenetic effects on human brain function? *Trends Cogn. Sci.* **19**, 366–373 (2015).
50. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10**, 1024–1032 (2015).
51. Derks, M. F. L. et al. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genom.* **17**, 332 (2016).
52. Cullinan, W. E., Herman, J. P., Battaglia, D. F., Akil, H. & Watson, S. J. Pattern and time course of immediate early gene expression in rat brain following acute stress. *Neuroscience* **64**, 477–505 (1995).
53. Aubin-Horth, N. & Renn, S. C. P. Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Mol. Ecol.* **18**, 3763–3780 (2009).
54. Bukhari, S. A. et al. Temporal dynamics of neurogenomic plasticity in response to social interactions in male threespined sticklebacks. *PLoS Genet.* **13**, e1006840 (2017).
55. Bell, A. M., Bukhari, S. A. & Sanogo, Y. O. Natural variation in brain gene expression profiles of aggressive and nonaggressive individual sticklebacks. *Behaviour* **153**, 1723–1743 (2016).
56. Trucchi, E. et al. BsRADseq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* **25**, 1697–1713 (2016).
57. Glastad, K. M., Gokhale, K., Liebig, J. & Goodisman, M. A. D. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 37110 (2016).
58. Verhulst, E. C. et al. Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD 4 DNA methylation. *Mol. Ecol.* **8**, 1801–1811 (2016).
59. Laine, V. N. et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474 (2016).
60. Cronican, A. A. et al. Genome-wide alteration of histone H3K9 acetylation pattern in mouse offspring prenatally exposed to arsenic. *PLoS ONE* **8**, e53478 (2013).
61. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
62. Hunter, C. P. Genetics: a touch of elegance with RNAi. *Curr. Biol.* **9**, R440–R442 (1999).
63. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
64. Straub, C., Granger, A. J., Saulnier, J. L. & Sabatini, B. L. CRISPR/Cas9-mediated gene knock-down in post-mitotic neurons. *PLoS ONE* **9**, e105584 (2014).
65. Swiech, L. et al. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106 (2015).
66. Peng, R., Lin, G. & Li, J. Potential pitfalls of CRISPR/Cas9-mediated genome editing. *FEBS J.* **283**, 1218–1231 (2016).
67. Chandrasekaran, S. et al. Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc. Natl Acad. Sci. USA* **108**, 18020–18025 (2011).
68. Wang, J. et al. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
69. Reidenbach, K. R. et al. Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in west and central Africa. *Genome Biol. Evol.* **4**, 1202–1212 (2012).
70. Lawniczak, M. K. N. et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330**, 512–514 (2010).
71. Chalfin, L. et al. Mapping ecologically relevant social behaviours by gene knockout in wild mice. *Nat. Commun.* **5**, 4569 (2014).
72. Jandt, J. M. et al. Behavioural syndromes and social insects: personality at multiple levels. *Biol. Rev.* **89**, 48–67 (2014).
73. Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**, 2728–2732 (2014).
74. Rausher, M. D. & Delph, L. F. Commentary: when does understanding phenotypic evolution require identification of the underlying genes? *Evolution* **69**, 1655–1664 (2015).
75. Lang, G. I., Murray, A. W. & Botstein, D. The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl Acad. Sci. USA* **106**, 5755–5760 (2009).
76. Cash, A. C., Whitfield, C. W., Ismail, N. & Robinson, G. E. Behavior and the limits of genomic plasticity: power and replicability in microarray analysis of honeybee brains. *Genes Brain Behav.* **4**, 267–271 (2005).
77. Zayed, A. & Robinson, G. E. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annu. Rev. Genet.* **46**, 591–615 (2012).
78. Cardoso, S. D., Teles, M. C. & Oliveira, R. F. Neurogenomic mechanisms of social plasticity. *J. Exp. Biol.* **218**, 140–149 (2015).

79. Réale, D. et al. Personality and the emergence of the pace-of-life syndrome concept at the population level. *Phil. Trans. R. Soc. B* **365**, 4051–4063 (2010).
80. Sanogo, Y. O. & Bell, A. M. Molecular mechanisms and the conflict between courtship and aggression in three-spined sticklebacks. *Mol. Ecol.* **25**, 4368–4376 (2016).
81. Zinzow-Kramer, W. M. et al. Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes Brain Behav.* **14**, 641–654 (2015).
82. Gibson, G. & Muse, S. V. *A Primer of Genome Science*. (Sinauer Associates: Sunderland, 2009).
83. Flicek, P. & Birney, E. Sense from sequence reads: methods for alignment and assembly. *Nat. Methods* **6**, S6–S12 (2009).
84. McGary, K. L. et al. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc. Natl Acad. Sci. USA* **107**, 6544–6549 (2010).
85. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
86. Baker, E. J., Jay, J. J., Bubier, J. A., Langston, M. A. & Chesler, E. J. GeneWeaver: a web-based system for integrative functional genomics. *Nucleic Acids Res.* **40**, D1067–D1076 (2012).
87. Rittschof, C. C. & Robinson, G. E. Genomics: moving behavioural ecology beyond the phenotypic gambit. *Anim. Behav.* **92**, 263–270 (2014).
88. Biro, P. A. & Stamps, J. A. Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends Ecol. Evol.* **25**, 653–659 (2010).
89. Biro, P. A. & Stamps, J. A. Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**, 361–368 (2008).
90. Sih, A. & Giudice, M. D. Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Phil. Trans. R. Soc. B* **367**, 2762–2772 (2012).
91. Tieleman, B. I., Williams, J. B., Ricklefs, R. E. & Klasing, K. C. Constitutive innate immunity is a component of the pace-of-life syndrome in tropical birds. *Proc. R. Soc. B* **272**, 1715–1720 (2005).
92. O'Connell, L. A. & Hofmann, H. A. The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. *J. Comp. Neurol.* **519**, 3599–3639 (2011).
93. Goodson, J. L. The vertebrate social behavior network: evolutionary themes and variations. *Horm. Behav.* **48**, 11–22 (2005).
94. Newman, S. W. The medial extended amygdala in male reproductive behavior. A node in the mammalian social behavior network. *Ann. NY Acad. Sci.* **877**, 242–257 (1999).
95. Newman, M. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
96. Newman, M. E. J. & Clauset, A. Structure and inference in annotated networks. *Nat. Commun.* **7**, 11863 (2016).
97. Wong-Riley, M. Changes in the visual system of monocularly sutured or enucleated cats demonstrable with cytochrome oxidase histochemistry. *Brain Res.* **171**, 11–28 (1979).
98. Laiho, J. E. et al. Relative sensitivity of immunohistochemistry, multiple reaction monitoring mass spectrometry, in situ hybridization and PCR to detect Coxsackievirus B1 in A549 cells. *J. Clin. Virol.* **77**, 21–28 (2016).
99. Knight, Z. A. et al. Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* **151**, 1126–1137 (2012).
100. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
101. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
102. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
103. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
104. Ketting, R. F. The many faces of RNAi. *Dev. Cell* **20**, 148–161 (2011).
105. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
106. Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247.e17 (2009).
107. Katz, P. S. 'Model organisms' in the light of evolution. *Curr. Biol.* **26**, R649–R650 (2016).
108. Saltz, J. B. Genetic composition of social groups influences male aggressive behaviour and fitness in natural genotypes of *Drosophila melanogaster*. *Proc. R. Soc. B* **280**, 20131926 (2013).
109. Egan, R. J. et al. Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. *Behav. Brain Res.* **205**, 38–44 (2009).
110. Amdam, G. V. & Page, R. E. Jr The developmental genetics and physiology of honeybee societies. *Anim. Behav.* **79**, 973–980 (2010).
111. Simola, D. F. et al. Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* **351**, aac6633 (2016).

Acknowledgements

The workshop that led to this series of papers was funded by the National Science Foundation (NSF-IOS 1623898; PI: A.M.B.), the NSF Sociogenomics Research Coordination Network and the Carl R. Woese Institute for Genomic Biology at the University of Illinois Urbana Champaign. We wish to thank other workshop participants for their feedback in the development of these ideas and comments on drafts of the manuscript.

Author contributions

S.E.B. contributed to conception of the manuscript, drafted sections of the manuscript, contributed to the conceptualization and generation of figures and tables, edited the manuscript and facilitated the collaboration between authors. R.A.D. drafted sections of the manuscript, contributed to the conceptualization and generation of figures and tables and provided feedback. Z.D. contributed to conception of the manuscript, drafted sections of the manuscript and provided feedback. S.M.P. contributed to conception of the manuscript, drafted sections of the manuscript and provided feedback. K.v.O. contributed to conception of the manuscript and provided feedback. A.S. contributed to conception of the manuscript and provided feedback. A.M.B. contributed to the conception of the manuscript, drafted sections of the manuscript, provided feedback and edited the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.