MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 725–739 doi: 10.1111/j.1365-294X.2010.04970.x

Allele-specific gene expression in a wild nonhuman primate population

J. TUNG,*† M. Y. AKINYI,‡§ S. MUTURA,‡ J. ALTMANN,‡¶ G. A. WRAY*†**[1] and S. C. ALBERTS*†‡**[1]

*Department of Biology, Duke University, PO Box 90338, Durham, NC 27708, USA, †Institute for Genome Sciences & Policy, Durham, NC 27708, USA, ‡Institute of Primate Research, National Museums of Kenya, PO Box 24481, Nairobi, Kenya, §Department of Medical Physiology, University of Nairobi, Nairobi, Kenya, ¶Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA, **Department of Evolutionary Anthropology, Duke University, PO Box 90383, Durham, NC 27708, USA

Abstract

Natural populations hold enormous potential for evolutionary genetic studies, especially when phenotypic, genetic and environmental data are all available on the same individuals. However, untangling the genotype-phenotype relationship in natural populations remains a major challenge. Here, we describe results of an investigation of one class of phenotype, allele-specific gene expression (ASGE), in the well-studied natural population of baboons of the Amboseli basin, Kenya. ASGE measurements identify cases in which one allele of a gene is overexpressed relative to the alternative allele of the same gene, *within* individuals, thus providing a control for background genetic and environmental effects. Here, we characterize the incidence of ASGE in the Amboseli baboon population, focusing on the genetic and environmental contributions to ASGE in a set of eleven genes involved in immunity and defence. Within this set, we identify evidence for common ASGE in four genes. We also present examples of two relationships between *cis*-regulatory genetic variants and the ASGE phenotype. Finally, we identify one case in which this relationship is influenced by a novel gene–environment interaction. Specifically, the dominance rank of an individual's mother during its early life (an aspect of that individual's social environment) influences the expression of the gene *CCL5* via an interaction with *cis*-regulatory genetic variation. These results illustrate how environmental and ecological data can be integrated into evolutionary genetic studies of functional variation in natural populations. They also highlight the potential importance of early life environmental variation in shaping the genetic architecture of complex traits in wild mammals.

Keywords: allele-specific gene expression, allelic imbalance, Amboseli baboons, *cis*-regulation, gene–environment interaction

Received 19 April 2010; revision received 6 November 2010; accepted 23 November 2010

Introduction

The relationship between genetic variation and phenotypic variation has a fundamental influence on evolutionary change. However, dissecting this relationship for traits of ecological and evolutionary relevance continues to be a substantial challenge. This is especially true in studies of nonmodel systems in the wild, for which inbred lines cannot be constructed and for which extensive genomic resources are not yet available. Nevertheless, some of the most ecologically and evolutionarily well characterized systems on the phenotypic level fall in this category (e.g. Grant 1986; Clutton-Brock 1989; Clutton-Brock & Pemberton 2004; Kruuk & Hill 2008; Tung *et al.* 2010). In these cases, prior knowledge about trait variation and its fitness impact in the wild

Correspondence: Jenny Tung, Fax: (773) 834 8470; E-mail: jtung@uchicago.edu

[1]These authors contributed equally to this work.

would doubly reward efforts to link genetic variation to phenotypic variation. For instance, while the finding that variation in the calmodulin and *BMP4* genes influences beak shape in Darwin's finches (*Geospiza* sp.) was itself a major contribution to evolutionary genetics (Abzhanov *et al.* 2004, 2006), its significance was greatly enhanced by the existence of long-term observational research on the relationship between beak morphology, feeding behaviour, and ecological niche differentiation (Grant 1986).

In general, the tools available for studying functional genetic variation in wild populations are more limited than for model systems investigated under laboratory conditions. However, allele-specific gene expression (ASGE) assays, also known as allelic imbalance assays, represent one potential strategy. These assays measure the relative contribution of the two alleles of the same gene to the total amount of mRNA for that gene *within* the same individual. When one allele drives significantly higher expression than the other allele, that gene shows evidence of ASGE. Thus, unlike most types of gene expression measurements, ASGE assays unambiguously indicate a causal basis for gene expression variation that lies in *cis* to the gene (a *cis*-acting effect influences only the copy of the gene on the same physical chromosome; effects that influence both alleles of the gene, such as environmental or genetic background effects, are *trans*-acting). This is because, in comparisons between alleles of a gene within individuals, the *trans* genetic and *trans* environmental backgrounds are held constant. Promisingly, ASGE assays can be broadly applied both to laboratory model systems (Wittkopp *et al.* 2004, 2008; de Meaux *et al.* 2005, 2006; Campbell *et al.* 2008; Gruber & Long 2009; Zhang & Borevitz 2009) and to nonmodel systems (Yan *et al.* 2002; Morley *et al.* 2004; Pastinen & Hudson 2004; Schaart *et al.* 2005; Cheung *et al.* 2008; Serre *et al.* 2008; von Korff *et al.* 2009; Tung *et al.* 2009; Heap *et al.* 2010), including from samples obtained directly from organisms in the field (Tung *et al.* 2009). Thus, unlike in vitro assessments of functional genetic variation, the functional effects of this variation in organisms experiencing natural environmental conditions are clear.

Allele-specific gene expression (ASGE) measurements have been used as a tool to disentangle the global contributions of *cis*- and *trans*-acting factors to gene regulation (Yan *et al.* 2002; Morley *et al.* 2004; Pastinen & Hudson 2004; Zhang & Borevitz 2009), to investigate the relationship between *cis*-regulatory variation and genetic divergence within and between species (Wittkopp *et al.* 2004, 2008; Gruber & Long 2009), and to test specific hypotheses about functional genetic variation that influences a given locus (de Meaux *et al.* 2005; Tao *et al.* 2006; Zhu *et al.* 2006; Linnen *et al.* 2009; Tung *et al.* 2009; Wittkopp *et al.* 2009; Babbitt *et al.* 2010). Most of this work has been conducted either in humans or in laboratory systems in which controlled crosses can be made, and ASGE assays have generally not been brought to bear in studies of natural populations (but see Linnen *et al.* 2009; Tung *et al.* 2009). However, the ability of ASGE measurements to control for *trans*-acting variation is particularly useful in work on natural populations, for which background effects and sampling conditions often cannot be standardized.

Here, we present an analysis of ASGE in a wild population of yellow baboons (*Papio cynocephalus*) that has been monitored continuously since 1971 as the focus of a long-term study in the Amboseli basin of southern Kenya (Altmann & Altmann 1970; Altmann *et al.* 1996; Buchan *et al.* 2003; Alberts *et al.* 2006). This research has produced a large body of knowledge on environmental and phenotypic variation in this population (Altmann & Alberts 2003; Silk *et al.* 2003; Beehner *et al.* 2006b; Charpentier *et al.* 2008a,b), making it an ideal candidate system for integrating genetic data into an existing ecological framework. The study subjects were members of a natural, unmanaged, unprovisioned population that receives no veterinary or other intervention and co-occurs with a full complement of natural predators.

We analysed expression data on ten genes associated with immune function and expressed in whole blood, and reanalyzed data on one gene (*FY*) that we previously characterized in another study (Tung *et al.* 2009), for a total gene set of eleven genes (Table 1). In addition, we took advantage of the long-term behavioural and environmental data for the individuals in our data set to extend our understanding of ASGE to include possible environmental interaction effects. Specifically, for two of these genes, we tested the hypothesis that variation in an individual's early social environment may have long-term effects on gene expression throughout life via gene–environment interactions (GEIs). Early environmental effects and GEIs involving the early environment have been demonstrated for a broad range of organisms and traits (e.g. Qvarnstrom 1999; David *et al.* 2000; Hoffjan *et al.* 2005), including in studies that indicate the importance of these interactions for immune genes (several interleukins and *IFNG*: Hoffjan *et al.* 2005; *CD14* and interleukin receptor 4-alpha: Suzuki *et al.* 2009). Given that gene expression often serves as a molecular precursor for such traits, we reasoned that early life environment might also have a long-reaching impact on expression.

We drew on knowledge from previous studies of the Amboseli baboons to identify a specific early life effect, maternal dominance rank, which might play a role in mediating GEIs on gene expression phenotypes.

Table 1 Genes included in this study

Gene	Gene name	Role	*n*	ASGE range*	*P*-value†
CCL5	Chemokine (CC motif) ligand 5	Pro-inflammatory chemokine	36	0.201–3.32	<0.0001
CCR5	Chemokine (CC motif) receptor 5	Membrane-bound chemokine receptor; T-cell entry point for HIV	25	−0.960–0.518	0.1651
CD14	Monocyte differentiation antigen CD14	Monocyte cell surface marker; recognizes bacterial lipopolysaccharide	7	0.112–0.425	0.0923
CXCR4	Chemokine (CXC motif) receptor 4	Membrane-bound chemokine receptor; T cell entry point for HIV	50	−0.420–0.418	0.0001
FY	Duffy antigen receptor for chemokines	Non-specific chemokine receptor; erythrocyte receptor for *Plasmodium vivax* malaria	38	−0.002–2.13	<0.0001
IL10	Interleukin 10	Anti-inflammatory cytokine	31	−0.491–0.108	<0.0001
IL1B	Interleukin 1-beta	Pro-inflammatory cytokine	33	−0.111–0.104	0.7103
IL6	Interleukin 6	Pro-inflammatory/anti-inflammatory cytokine	13	−0.741–0.001	0.0011
LTA	Lymphotoxin alpha	Lymphocytic cytokine involved in the inflammatory and antiviral response	36	−0.429–0.463	0.5798
TAP2	Transporter, ATP binding cassette, major histocompatibility complex, 2	MHC cluster gene involved in antigen presentation to T cells	15	−0.481–0.402	0.4512
TNF	Tumour necrosis factor	Pro-inflammatory cytokine, also involved in apoptosis	8	−0.125–0.039	0.0781

ASGE, allele-specific gene expression

*Range refers to the range of mean $\log_2$-transformed corrected ASGE values for each individual, across all replicate measurements. CCL5, CXCR4, FY, and IL10 reflect samples collected from 2005 to 2009; all other genes include samples from 2005 to 2008.

†Uncorrected *P*-values for common ASGE were derived from 10 000 random permutations of the data, as described in Materials and methods.

Although maternal dominance rank, defined as the dominance rank (i.e. social status) of an individual's mother at the time of that individual's conception, is a trait of the mother, it also represents an important environmental variable in the early life of her offspring: like humans, infant baboons in the periparturitional period are completely dependent on their mothers for resource acquisition and protection, and a mother's social status influences the quality of this environment, sometimes with long-term effects. For instance, maternal dominance rank in the Amboseli baboons predicts the glucocorticoid levels of male offspring during late adolescence more effectively than a male's own rank, years after the environmental effect was experienced, and even in cases in which the mother has previously died (Onyango *et al.* 2008). Additionally, maternal dominance rank affects offspring growth rates (Johnson 2003; Altmann & Alberts 2005) and the age that offspring mature (Altmann & Alberts 2005; Charpentier *et al.* 2008a). Further, maternal social connectedness during the early life of her offspring predicts offspring survival in the first year of life (Silk *et al.* 2003), an effect that has been replicated and extended to encompass overall lifetime survival in the chacma baboons (*Papio ursinus*) of the Okavango Delta (Silk *et al.* 2009).

By combining allele-specific expression assays, genotype data, and field observations, we therefore sought to integrate environmental effects, genetic effects, and phenotypic data to understand the architecture of gene expression in a natural population. In particular, we aimed to provide a first overview of the role of ASGE in the Amboseli baboons, including its link to genetic variation in nearby *cis*-regulatory regions and its relationship to maternal dominance rank, an important early life environmental effect.

Materials and methods

Study subjects

The Amboseli basin is a semi-arid short-grass savanna in southern Kenya, bordering Tanzania on the south. The Amboseli baboon population consists of primarily yellow baboons (*Papio cynocephalus*) with some hybrid admixture from immigration of anubis baboons (*Papio anubis*) from outside the basin (Samuels & Altmann 1986; Alberts & Altmann 2001; Tung *et al.* 2008). Five study groups composed of individually recognized animals are currently monitored on a near-daily basis within the larger population: life history, behavioural,

and physiological data are recorded for all individuals, maternal pedigrees are available for all natal individuals, and paternal pedigrees are available for many (Buchan *et al.* 2003; Alberts *et al.* 2006). The subjects of this study were 101 adult baboons (55 females and 46 males), representing approximately 60% of the adults in the study population. Study subjects lived in one of our five main study groups (monitored on a near-daily basis) or in one group that is monitored monthly for demographic information only. Study subjects were chosen based on an opportunistic darting strategy: on days pre-arranged as darting days, we targeted animals that could be darted without being observed by their conspecifics (for instance, because other animals were momentarily oriented away from them). This resulted in darting choices that were random with respect to particular characteristics of the animals, with the important exception that we only darted adult (post-pubertal) animals, and we avoided darting females in the second or third trimesters of pregnancy and those with dependent infants. Samples were collected between 2005 and 2009.

All study subjects were anesthetized with an anaesthetic-bearing dart using a handheld blowgun. Darting occurred in the morning (0700–1200), when animals descended from known sleeping sites. In order to minimize disruption to the study groups, darting only occurred when no individuals within the group would observe the actual dart delivery, and we darted no more than two animals per day, no more than 3 days a week. Anesthetized baboons were quickly removed to a processing site distant from the rest of the group. We collected RNA samples by drawing two 2.5 mL samples of whole blood into PaxGene Vacutainer tubes (BD Vacutainer), which protect RNA from environmental degradation and prevent further transcription after the blood draw. We also collected blood samples for DNA extraction. Upon regaining consciousness, study subjects were placed into a covered holding cage until fully recovered from the effects of the anaesthetic (~3–4 h). They were then released in the vicinity of their social group. All subjects rejoined their social groups quickly upon release and without incident.

Blood samples were stored for no more than 3 days in an evaporatively-cooled charcoal structure at Amboseli, which maintains a temperature of 20–25 °C. They were then shipped to Nairobi, where they were either preserved frozen at −20 °C until they could be hand couriered to the United States, or, in a subset of cases, immediately extracted at the Institute of Primate Research in Nairobi (see Appendix S1 in Supporting information). RNA extractions were conducted using the PaxGene RNA kit (Qiagen) and RNA was reverse transcribed into cDNA (High Capacity cDNA Archive kit; Applied Biosystems) for subsequent pyrosequencing. The robustness of the ASGE measurements to these protocols was confirmed by systematic comparisons between samples extracted from the same individual after different storage and transport conditions (Table S1 and Fig. S1, Supporting information). DNA samples were extracted for each study subject using the DNEasy DNA Extraction kit (Qiagen).

Candidate gene assay development

All eleven candidate loci used in this study are well studied in humans with respect to disease risk and progression, and all contain segregating genetic variants in human populations that have been specifically associated with disease-related phenotypes, many of which are *cis*-regulatory (McKusick-Nathans Institute of Genetic Medicine & National Center for Biotechnology Information 2010). All the genes in the data set were also linked by their involvement in immune function, reflecting one of the main biological functions of the tissue from which we sampled RNA. Additionally, either intraspecific sequence data or interspecific comparisons in humans or nonhuman primates have suggested interesting selective patterns for several of these loci (for example, *CCR5*: Bamshad *et al.* 2002; *FY*: Hamblin & Di Rienzo 2000; Hamblin *et al.* 2002; *CCL5*, *IL10*, *IL1B*, *LTA*: Hughes *et al.* 2005).

All methods of measuring ASGE depend on the presence of one or more segregating SNP variants in the transcribed region of a target gene. Allele-specific assays are applied to individuals who are heterozygous for this variant, because by discriminating between the two alleles at the transcribed SNP, the two gene transcripts (and, by proxy, their linked *cis*-regulatory regions) can also be differentiated. We identified common transcribed SNPs segregating in the Amboseli baboon population by sequencing transcribed regions of the 11 candidate loci in an ascertainment panel of 10–12 unrelated baboons. We focused specifically on identifying intermediate frequency SNPs. These SNPs are the most useful variants for constructing ASGE assays because multiple individuals are likely to be heterozygous at these sites. Thus, we designed ASGE assays only around transcribed SNPs with an estimated minor allele frequency of at least 10%. We identified suitable transcribed SNPs for all eleven loci, and designed one assay each for all genes except for *FY*, for which we designed two assays as reported in Tung *et al.* (2009). Although we have already reported some of the results for the *FY* gene in previous work, we included it here for the purpose of comparison to the other ten loci, and also in order to explore possible GEIs influencing ASGE at *FY* (see below).

We then genotyped these SNPs using pyrosequencing (using the PyroMark Q96 MD instrument and PyroGold reagents, Biotage) or direct sequencing (using an ABI 3730×l sequencer and Big Dye Terminator reagents, version 3.1; Applied Biosystems) for all individuals sampled from 2005 to 2007. Approximately 1.05% of the genotypes is missing in this dataset as a result of failed genotyping or sequencing reactions. A subset of the genes exhibited common allelic imbalance within the samples collected between 2005–2007, based on comparisons between genomic DNA and cDNA measurements (see below for further details). For these genes, we also genotyped and assayed individuals sampled between 2008 and 2009 (0.09% missing genotypes in this total set). At least seven heterozygotes (range: 7–37 heterozygotes per gene, mean: 25.1) were assayed for each of the genes in this study (Table 1).

All sequences were visually inspected for ascertainment of variable sites in the population and identification of heterozygous individuals using Sequencher 4.8 (GeneCodes). Pyrosequencing genotypes were assigned by calculation of relative peak heights at the variable site and/or by automated assignment using PyroMark MD software.

ASGE measurements via pyrosequencing

Allele-specific expression assays were conducted using pyrosequencing on a PyroMark Q96 MD instrument. Briefly, pyrosequencing is a genotyping approach based on cycle sequencing, in which the successful addition of a complementary base to a sequencing template produces light emissions. When the template contains a heterozygous SNP, as in ASGE assays, the amount of light produced upon addition of one complement vs. the alternative complement at that SNP reflects the relative prevalence of the two templates (e.g. Yan *et al.* 2002; Wittkopp *et al.* 2004). We conducted pyrosequencing-based ASGE assays for individuals that were heterozygous at the transcribed assay SNP for each candidate locus (this SNP usually occurs at a silent site or in an untranslated region of the gene, and thus differentiates RNA transcripts, but not the resulting protein). We generated template for the ASGE assay by running PCRs on the cDNA produced from the samples drawn in PaxGene tubes or on genomic DNA. For each gene-individual combination, we ran four replicates produced from four independent initial PCRs on each of two replicate plates. Thus, a total of eight gene expression measurements were obtained for each individual for each candidate gene. For *IL6*, greater technical variance in the assay led us to measure each individual twelve times (across three plates). Each measurement corresponds to the ratio of expression of one allele of the gene vs. the alternative allele of the same gene. For example, if two alleles could be discriminated based on a C/T transcribed SNP, allele-specific differences in expression would be represented as the signal for the allele carrying the 'C' variant divided by the signal for the allele carrying the 'T' variant, corrected by any technical bias in relative signal strength detected from concurrent measurements on genomic DNA for the same individual (see Appendix S1 in Supporting information). Corrected ratios were then $\log_2$-transformed for downstream analyses.

Assessment of ASGE for each locus

In order to identify functional *cis*-regulatory variants in the Amboseli population, we focused on genes that commonly exhibited ASGE within our sample. Rare cases of ASGE are more likely to reflect rare genetic variants, rare interactions with environmental or genetic background effects, or the presence of multiple functional variants with opposing effects. These situations make it difficult to associate ASGE with *cis*-regulatory genetic variation, and so we avoided them. We identified cases of common ASGE by comparing the measurements made on cDNA for a given gene to the control genomic DNA (gDNA) measurements for the same gene, using raw $\log_2$-transformed values for both sets of measurements. We considered ASGE to be common for a gene when the distribution of allelic imbalance for that gene was significantly different for cDNA samples than the corresponding distribution for gDNA samples, over all individuals. Note that this approach does not identify significant allelic imbalance on an individual-by-individual basis, but instead tests for a shift away from equal expression of both alleles in the mean for the population (in all subsequent analyses focused on individual differences, we treat ASGE levels as a continuous trait). This approach eliminated genes for which ASGE was rare (i.e. affecting one or only a few individuals in the sample), because these individuals did not strongly influence the overall distribution of ASGE values.

We evaluated the significance of common ASGE for each gene by randomly permuting the labels (cDNA or gDNA) over these values. Because more cDNA measurements were made for each individual than gDNA measurements, we first randomly subsampled the number of cDNA measurements for each individual to equal the number of gDNA measurements. We repeated this subsampling routine 10 000 times. We then calculated a *P*-value for each subsampled data set using a two-tailed nonparametric Wilcoxon summed ranks test, which tested whether the cDNA values were significantly different than the values for the gDNA set. We took the

mean of this set of *P*-values to be the nominal *P*-value for the gene. This value was then compared to a distribution of *P*-values derived from random permutations, which provided a null distribution on *P*-values for each gene.

Sequencing of gene regulatory regions

Four genes exhibited strong evidence for common allelic imbalance (*CCL5*, *FY*, *CXCR4*, *IL10*; see Results). In order to identify genetic variants associated with ASGE in these genes, we sequenced 0.65–0.82 kilobases upstream of the transcription start site for the set of individuals assayed for each of these genes (Tables S6–S9, Supporting information). For *IL10*, we also sequenced 0.72 kilobases in the 3′ untranslated region and 3′ flanking region because our analyses suggested that the upstream sequence did not explain observed ASGE patterns and because the assay SNP used for this gene is located in its last exon (which also contains the 3′ UTR; Table S10, Supporting information). Variable sites were identified by visual examination of the resulting sequence traces, and genotype assignments were produced for each individual-gene combination based on the sequence data.

Association between ASGE data and regulatory variants

ASGE is a gene expression phenotype caused by *cis*-regulatory genetic variants that functionally differ in their abilities to drive gene expression. Because ASGE reflects the ratio of gene expression between alleles within individuals, only individuals that are heterozygous at one or more of these variants will exhibit significant ASGE. This leads to the expectation that heterozygotes at a functional *cis*-regulatory variant will exhibit more extreme values of ASGE than homozygotes for the same variant. We used this expectation to test for an association between the *cis*-regulatory variants detected in the sequenced regulatory regions of the four commonly imbalanced genes, and the ASGE data obtained from the pyrosequencing assays. For these genes, we expanded the original dataset (individuals darted during 2005–2007) to include an additional set of 32 individuals darted during 2008–2009, as indicated above.

Testing for an association between ASGE and genotype at a putative regulatory SNP is possible because, although individuals assayed for ASGE are always heterozygous at the pyrosequencing assay SNP, they may be either heterozygous or homozygous at a SNP in the *cis*-regulatory region due to historic recombination. Thus, an individual who is heterozygous at the transcribed assay SNP could be heterozygous at an upstream regulatory SNP, or homozygous at that regulatory site, implying incomplete linkage disequilibrium between these sites. Indeed, variation in the level of ASGE among individuals that are all heterozygous at the assay SNP suggests that this case often holds true.

We used general linear mixed models to analyse variation in allelic imbalance in all assayed individuals. Genotypes were coded as heterozygous or homozygous at known *cis*-regulatory variable sites (Table 2; excluding singletons in the sample, which are too rare to account for common ASGE and also impossible to analyse in this context) and treated as fixed effects within the model. In cases where two *cis*-regulatory SNPs were perfectly correlated (Tables S6–S10, Supporting information), we analysed genotype at only one representative site. In total, we investigated three sites for *CCL5*; one site for *CXCR4*; four sites for *FY*; and six sites for *IL10*. Year of sampling was treated as a random effect (Tables S2–S5, Supporting information). Parameter estimates for all model effects were conducted using the *lme4* package in R 2.8.1 (R Development Core Team 2007). For two genes, *CXCR4* and *IL10*, the distribution of ASGE included a large number of both negative and

Table 2 Coefficients, *P*-values and sample distribution for genotype effects in the fit models for genes with significant common allele-specific gene expression in the Amboseli population (bold values represent significant effects).

Gene	SNP	# Hets, # Homs*	Coefficient†	*P*-value
CCL5	SNP1 (C/T)	14 CT, 22 TT	0.463	0.879
	SNP2 (A/G)	6 AG, 30 AA	−0.278	0.154
	SNP3 (A/G)	14 AG, 22 GG	−1.787	**<0.0001**
CXCR4	SNP1 (A/G)	17 AG, 19 GG	0.065	0.956
FY	SNP1 (A/G)	29 AG, 8 AA	0.059	0.572
	SNP2 (A/G)	2 AG, 36 GG	0.277	0.698
	SNP3 (A/G)	4 AG, 34 AA	0.235	0.738
	SNP4 (C/T)	18 CT, 20 CC	−0.620	**0.0002**
IL10‡	SNP1 (C/T)	4 CT, 23 CC	−0.035	0.176
	SNP2 (C/T)	2 CT, 24 CC	−0.022	0.415
	SNP3 (C/T)	2 CT, 21 CC	−0.096	0.082
	SNP4 (C/T)	6 CT, 18 CC	0.035	0.604
	SNP5 (C/G)	16 CG, 8 GG	−0.024	0.411
	SNP6 (C/T)	8 CT, 17 CC	0.042	0.756

*The sum of each cell is equivalent to the number of individuals with genotype information at that site. For some individuals, genotype data were not included due to missing or poor quality sequence.

†A negative coefficient corresponds to increased levels of allelic imbalance in heterozygotes vs. homozygotes.

‡SNP1 and SNP2 are located in the putative upstream *cis*-regulatory region for *IL10*; SNPs 3 – 6 are located in the 3′ untranslated region and 3′ flanking region. Additional information about the position of each SNP is provided in Tables S6–S10, Supporting information.

positive values (i.e. imbalance was detected at the assay SNP in both directions; Table 1). This effect may result from low levels of linkage disequilibrium between an assay SNP and a causal *cis*-regulatory SNP, leading to a case in which heterozygotes at the regulatory site exhibit both more negative and more positive values of ASGE than homozygotes. To avoid misidentifying the genes that exhibited increased variance as genes for which *cis*-regulatory genotype has no effect, we therefore used 'unsigned' ASGE values (i.e. the absolute value of the $\log_2$-transformed ASGE values; see for example Babbitt *et al.* 2010) for these two genes.

We assigned *P*-values for each model effect using random permutations of the allelic imbalance measurements for a given individual against individual identity (as in Tung *et al.* 2009). We then used a backwards model selection procedure to sequentially eliminate the model effect with the highest *P*-value, until all *P*-values were below 0.10. We performed significance testing using permutations in order to standardize our methods across genes for which ASGE values followed different distributions (permutation testing is a non-parametric approach), and because this approach allowed us to retain all replicate measurements in the analysis, instead of using a summary value such as a mean or median (we permuted blocks of replicate measurements for each individual). For each gene, several individuals in the data set were close relatives; to ensure that genetic correlations between these individuals did not produce a false signal of association, we also analysed the data after eliminating individuals in the data set so that no close relatives were included. In each case, eliminating different sets of individuals could produce this outcome; however, in no case did elimination of close relatives qualitatively change the results.

GEI effects on gene expression

A significant correlation between an ASGE phenotype and environmental effects suggests the presence of a GEI in which the environment, acting in *trans* to influence both alleles of a gene, modifies the effect of the *cis*-regulatory variant(s) (de Meaux *et al.* 2005; von Korff *et al.* 2009). Understanding GEIs that influence evolution in the wild may prove to be a particularly important contribution of evolutionary genetic work on natural populations. To explore this possibility, we tested for the presence of GEIs involving *cis*-regulatory variation in the two genes for which significant common ASGE was detected and could be associated with a known *cis*-regulatory genotype. We focused on an environmental effect of known importance in the Amboseli population: the social rank of an individual's mother, at the time of that individual's conception (i.e. maternal dominance rank), which is known to exert long-term effects on maturation timing and stress hormone profiles in this population (Alberts & Altmann 1995; Charpentier *et al.* 2008a; Onyango *et al.* 2008). Data on maternal dominance rank was extracted from the long-term relational database for the Amboseli Baboon Research Project, BABASE. Ranks are calculated on a monthly basis from a matrix of agonistic interactions between adult females residing in the same social group (Altmann & Alberts 2005). A mother's rank at the time of conception is therefore her assigned rank for the month in which she conceived the focal individual in our data set. Conception dates are determined from near-daily reproductive records on each female that track stages of the menstrual cycle based on external morphological information, a technique validated by faecal hormones (Beehner *et al.* 2006a; Gesquiere *et al.* 2007).

To investigate the possibility of GEIs that influence ASGE, we analysed the residuals from the previous model of ASGE on genotype in the context of a general linear model (see also Appendix S1 in Supporting information). We stratified the data by genotypic class at the associated regulatory SNP (i.e. by whether the individual was heterozygous or homozygous at the regulatory SNP), which permitted us to analyse the relationship between maternal dominance rank and ASGE separately for homozygotes and heterozygotes. This stratification allowed us to test the expectation that, if an environmental effect modifies the effect of a functional *cis*-regulatory variant, a relationship between ASGE and the environment should be observed in heterozygotes for this variant, but not in homozygotes. This expectation arises because an environmental interaction with *cis*-regulatory variation should not be observable via ASGE measurements if the two alleles for an individual are not functionally differentiated (i.e. if the individual is homozygous). We conducted this analysis subsequent to the initial genotype-phenotype association analysis in order to specifically test the hypothesis that observed associations between genetic effects and ASGE are modified by the environmental effects of maternal dominance rank (as opposed to testing the hypothesis that an interaction exists between any SNP and maternal dominance rank, an approach that would be difficult to biologically interpret and that would involve fitting a much larger number of model parameters). Consequently, we restricted the GEI analysis to the two genes in our data set that showed evidence for an association between *cis*-regulatory genotype and ASGE, *CCL5* and *FY*.

The P-values for this analysis were assigned by running the same analysis after permuting the response variable (residuals of a model taking into account year

of sampling and the genotype effect) with respect to the explanatory environmental variable. A null distribution of effects was obtained from 1000 random permutations, and the probability of observing an effect size greater than the estimated effect from the unpermuted data was taken as the *P*-value for the test (equivalent to a two-tailed test).

Results

ASGE is common in the Amboseli baboons

Of the eleven loci included in this study, four (36.4%) exhibited significant common ASGE (i.e. significant sample set-wide differences between ASGE ratios derived from cDNA vs. ASGE ratios derived from genomic DNA) at $P < 0.01$ following Bonferroni correction for multiple testing (*CCL5*, *CXCR4*, *FY*, *IL10*; Table 1, Fig. 1).

Among these four genes, we detected a range of effect sizes. For example, for *CCL5*, cDNA measurements almost never overlapped with genomic DNA measurements for any individual, while for *CXCR4*, *FY* and *IL10*, the cDNA and gDNA distributions overlapped somewhat (Fig. 1). In addition, the average corrected, $\log_2$-transformed ASGE measurement for an individual assayed at *CCL5* was 1.38 (range: 0.201–3.32), corresponding to a fold change difference in gene expression for the higher expressing allele of 2.60. In contrast, the average corrected ASGE measurement for an individual assayed at *CXCR4* was 0.112 (range: −0.420 to 0.418; 1.08 fold change difference between alleles). Nonetheless, all four of these genes showed significant ASGE, in contrast to the other seven genes in our dataset, exemplified by *CCR5* and *LTA* in Fig. 1.

ASGE is associated with cis*-regulatory genetic variation*

Of the four genes exhibiting common ASGE (*CCL5*, *CXCR4*, *IL10* and *FY*), two (including *FY*, as previously reported in Tung *et al.* 2009) exhibited an association between heterozygosity/homozygosity at a putative *cis*-regulatory genetic variant and magnitude of ASGE, such that heterozygotes exhibited more extreme ASGE than homozygotes (Fig. 2). As in the case of ASGE itself, we observed considerable variation in the size of the effect of the putative regulatory variant (see also Appendix S1 in Supporting information). For *CCL5*, genotype at the associated variant (an A/G transition located 200 base pairs upstream of the translation start site) explained 66.5% of the variance in the overall set of ASGE measurements, after taking into account the effects of year of sampling ($P < 0.0001$, $n = 14$ heterozygotes and 22 homozygotes). Additionally, we observed no overlap between the range of ASGE for heterozygotes at this site (range of mean per individual $\log_2$-transformed ratios: 1.77–3.32) and the range of ASGE for homozygotes at this site (range of mean per individual $\log_2$-transformed ratios: 0.201–1.52). For *FY*, the variant associated with ASGE (a C/T transition located 674 base pairs upstream of the erythrocytic

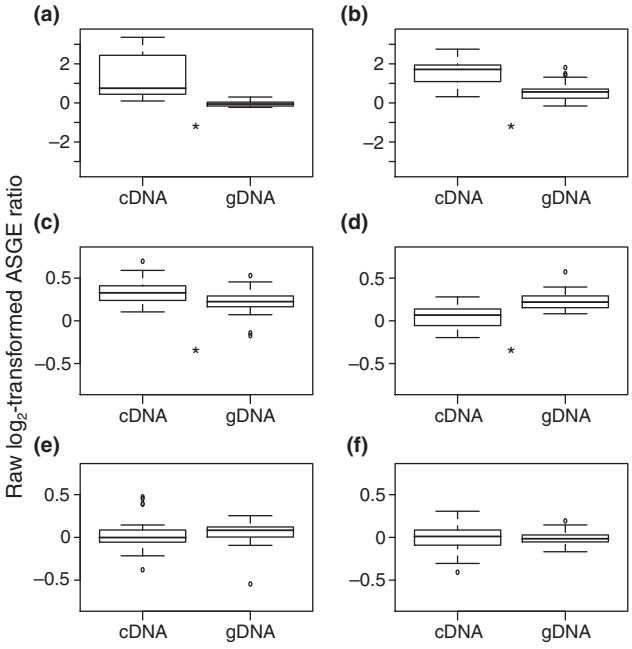

Fig. 1 Example allele-specific gene expression (ASGE) ratios for cDNA and genomic DNA (gDNA) for six genes. (a) *CCL5*; (b) *FY*; (c) *CXCR4*; and (d) *IL10* illustrate significant differences in $\log_2$-transformed ASGE ratios between cDNA samples and gDNA samples (indicated by an asterisk); whereas, for comparison, (e) *CCR5* and (f) *LTA* illustrate cases of statistically indistinguishable cDNA and gDNA measurements.

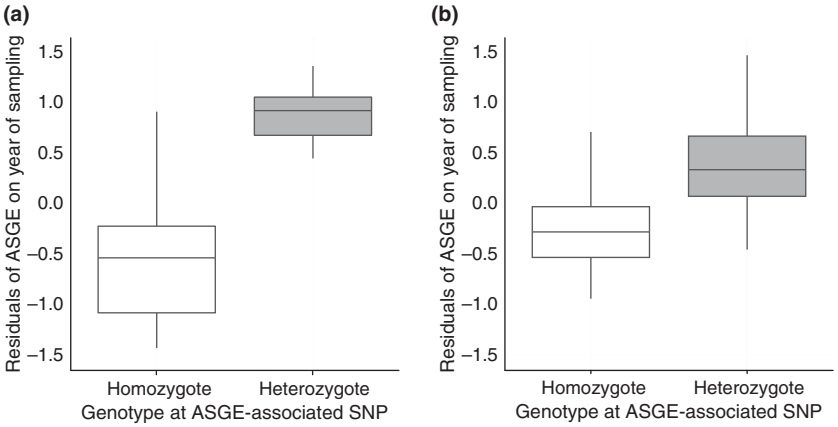

Fig. 2 Heterozygotes at allele-specific gene expression (ASGE)-associated SNPs exhibit more extreme levels of ASGE than homozygotes at (a) *CCL5* ($P < 0.0001$) and (b) *FY* ($P = 0.0002$). Although the year of sampling (i.e. batch) effect was estimated as a random effect within the full linear model, the *y*-axis here depicts the residuals of ASGE measurements after controlling for year of sampling, in order to highlight the genotype effect.

translation start site) explained a more modest proportion of the overall ASGE variance, 22% ($P = 0.0002$, $n = 18$ heterozygotes and 20 homozygotes), and the ranges for ASGE measurements overlapped between heterozygotes and homozygotes (heterozygotes: 0.203–2.131; homozygotes: −0.002 to 1.46).

For both *CCL5* and *FY*, the putative functional SNPs we identified were the closest SNPs in each set to the transcription start site. Interestingly, and also in both cases, our results suggest that additional functional variants and/or *cis*-by-*trans* regulatory interactions influence expression of these genes. In particular, for *CCL5*, even the homozygotes for the associated *cis*-regulatory variant exhibited strong signals of ASGE, although the magnitude of allelic imbalance for these animals was attenuated relative to heterozygotes.

For *CXCR4* and *IL10*, we were unable to identify an association between ASGE and SNP genotypes in the immediate *cis*-regulatory region. Both of these genes exhibited more modest levels of ASGE than *CCL5* and *FY* (Table 1) and were characterized by variation in ASGE that encompassed both positive $\log_2$-transformed ratios and negative $\log_2$-transformed ratios, probably reducing the power to detect an association. Additionally, we surveyed only a small region of sequence in which *cis*-regulatory variants may occur.

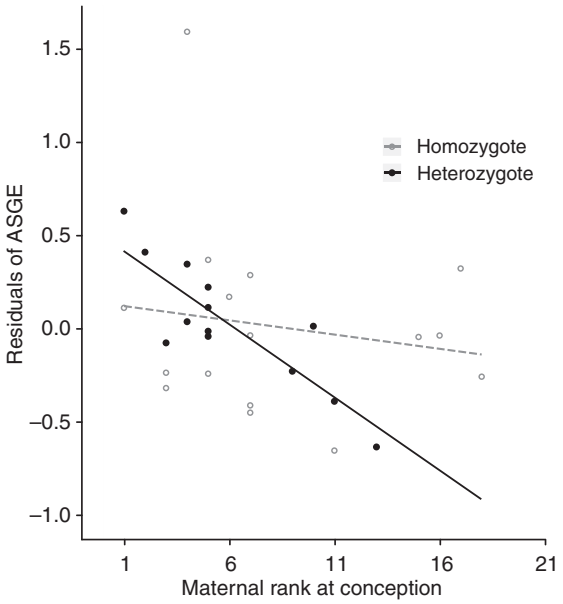

Fig. 3 Maternal rank at conception influences allelic imbalance in heterozygotes at the *CCL5* putative functional *cis*-regulatory site ($P < 0.001$, $R^2_{adj} = 0.706$), but not homozygotes ($P = 0.464$), after taking into account year of sampling and the direct effect of the allele-specific gene expression-associated SNP (total $n = 29$). High ranking individuals have low rank numbers (rank 1 is highest); low ranking individuals have high rank numbers.

GEI analysis

For *CCL5*, we identified an effect of maternal rank at the time of an individual's conception on gene expression, such that high maternal rank was correlated with more pronounced ASGE in individuals that were heterozygous for the putative functional *cis*-regulatory site, after controlling for the direct effect of genotype on ASGE and year of sampling ($P < 0.001$; Fig. 3). No such effect was detected for homozygotes at the *cis*-regulatory site ($P = 0.464$), as expected if maternal rank interacts with genetic variation captured by this site.

This result suggested that the early life environment influences gene expression at CCL5. However, if the early environment we considered, maternal dominance rank, is correlated with an individual's own rank in adulthood, the observed effect might instead represent an interaction between genetic variation and the current environment. Indeed, in baboons, maternal rank exerts a non-genetic effect on the dominance rank achieved by female offspring in adulthood (Walters 1987). We therefore checked whether the early life maternal rank effect

can be distinguished from an effect of the individual's own rank at the time of sampling (i.e. the individual's current environment). Although maternal rank was significantly correlated with rank at time of sampling in the *CCL5* data set (Spearman's rho = 0.492, $P = 0.008$, $n = 28$ individuals for which both ranks were known), we found no evidence that an individual's own rank at sampling influenced ASGE in *CCL5 cis*-regulatory variant heterozygotes ($P = 0.918$). Together, these results suggest a long-term effect of early life environment on *CCL5* expression, even in the absence of an effect of a similar environment at the time of sampling. In contrast, we found no evidence for GEI involving maternal rank on *FY* expression ($P = 0.766$).

Discussion

ASGE in the Amboseli baboon population

The presence of common ASGE was well supported for four of the 11 genes included in this study. Although it is difficult to compare rates of ASGE across studies, given different kinds and numbers of samples, different measurement platforms and different statistical methods, this frequency falls well within the rather broad range of previous estimates for different taxa given in the literature (from 5% to 70% in humans, mice, *Drosophila*, and *Arabidopsis*: Yan *et al.* 2002; Pant *et al.* 2006; Milani *et al.* 2007; Campbell *et al.* 2008; Serre *et al.* 2008; Gruber & Long 2009; Zhang & Borevitz 2009; Heap *et al.* 2010). This result suggests that our data set provides a good first glimpse into the functional *cis*-regulatory landscape of this population. However, because the genes in our data set all share a role in immune function, and because they were all chosen in part because of prior information on genetic variation in humans and other primates, further work on additional genes will be necessary in order to test whether these results typify the genome.

The four genes that exhibited evidence for common ASGE varied in both the range of ASGE detected across individuals and in our ability to map them onto genetic variants. Notably, the two cases of ASGE that we were able to link to *cis*-regulatory genetic variants were those that exhibited the greatest magnitude of ASGE, and the most consistent direction of imbalance. A greater magnitude of ASGE reflects more pronounced functional differentiation between *cis*-regulatory alleles, whereas a consistent direction for imbalance measurements indicates that the *cis*-regulatory variant(s) responsible for ASGE is relatively tightly linked to the transcribed region of the gene (and to the observed associated variants, if they themselves are not causal). These factors imply, perhaps unsurprisingly, that genes that exhibit high frequencies and magnitudes of allelic imbalance, along with consistent overexpression of the same allele, will be the most tractable subjects for studies of ASGE. Additionally, our results suggest that prior functional studies in model and/or laboratory organisms can inform the choice of candidate genes in natural populations, as exemplified by work incorporating ASGE measurements in *Drosophila* (Wittkopp *et al.* 2009), mice (Linnen *et al.* 2009), and primates (Wooding *et al.* 2006; Tung *et al.* 2009).

Our failure to associate ASGE with *cis*-regulatory variation at two loci that exhibit common allelic imbalance, *CXCR4* and *IL10*, highlights the fact that, in some cases, associating ASGE with putative causal regulatory variants will be difficult. This is particularly true for species like baboons, for which relatively little is known about segregating genetic variation. Indeed, in some cases, *cis*-regulatory variants actually occur many kilobases away from the transcribed sequence (reviewed in Wray *et al.* 2003), well outside the scope of surveys of sequence close to the gene. Thus, even if a distant causal variant is genotyped, it may not be in strong linkage disequilibrium with a transcribed assay SNP, weakening the power to detect a consistent relationship between genotype at a regulatory site and the ASGE phenotype. Similarly, strong interaction effects between two causal variants, or between a causal variant and an environmental effect, could also reduce power to detect such a relationship. Finally, ASGE could also reflect allele-specific epigenetic differences that lead to differential regulation between alleles in the absence of genetic variation in the *cis*-regulatory sequence. One or several of these conditions may have held in the cases of *CXCR4* and *IL10*. However, recent evidence from humans suggests that most functional *cis*-regulatory variants probably do lie close to either the transcription start site or the transcription end site (Veyrieras *et al.* 2008). Additionally, a genome-wide survey of allele-specific methylation patterns in humans suggests that up to 90% of the cases of allele-specific methylation are due to local sequence features acting in *cis* (i.e. *cis*-regulatory sequence variation) (Schalkwyk *et al.* 2010). Thus, rather than being mutually exclusive, epigenetic factors may provide a mechanism by which *cis*-regulatory genetic variants act.

Indeed, an expanded search space for functional *cis*-regulatory variants would likely also benefit the analysis even of genes like *CCL5* and *FY*, for which we identified significant associations between ASGE and *cis*-regulatory variation. In both cases, substantial variation in ASGE measurements remained unexplained by our analysis. In the case of *CCL5*, all homozygotes at the associated *cis*-regulatory SNP also exhibited strong evidence of allelic imbalance, although it was reduced relative to heterozygotes at this site. In agreement with

results from humans (Tao *et al.* 2006) and *Drosophila* (McGregor *et al.* 2007; Gruber & Long 2009), then, our findings suggest that ASGE is a complex trait influenced by multiple genetic variants, and that *cis*-regulatory interactions with *trans*-acting environmental or regulatory genetic effects may also play a role.

From a functional perspective, the four genes in this study that exhibited the best evidence for common ASGE in the Amboseli population (*CCL5*, *CXCR4*, *FY*, and *IL10*) are linked by their roles in the primate immune system. All four genes are involved in cytokine or chemokine signalling and play a part in mediating the inflammatory response. Controlling inflammation is a crucial component of the immune response, and is almost certainly important for wild baboons, which are subject to a wide array of both pathogen infections and physical insults; we have noted enlarged and inflamed lymph nodes on many of the baboons sampled for this study (unpublished data). Functional differences that vary the expression levels of these genes may therefore prove important in fine-tuning this response, and our results indicate that substantial genetic variation is segregating in the Amboseli baboons that could impact phenotypic evolution in this population.

Further, our results for *CCL5* suggest that at least some of this genetic variation may be influenced by GEIs, including environmental effects that are far removed in time from the gene expression measurement itself. Prior work in this population already implicated maternal dominance rank as a factor in long term phenotypic development (Onyango *et al.* 2008), and early life effects involving social status and access to resources are well known in humans and other animals (Ravelli *et al.* 1976; Altmann 1991; Lindstrom 1999; Qvarnstrom 1999; Godfrey & Barker 2000; Reifsnyder *et al.* 2000; Barker 2002; Barker *et al.* 2002; Weaver *et al.* 2004; Hoffjan *et al.* 2005; Meaney & Szyf 2005; St Clair *et al.* 2005), including some involving the effect of early life social status on gene expression (Miller *et al.* 2009). To our knowledge, however, these data represent the first evidence of an early life effect interacting with genotype in wild primates. With respect to this specific case, they suggest the possibility that early social environment in baboons may help shape the baboon immune system over the long-term, such that the same environmental exposure produces different consequences for different individuals. This result echoes findings in humans that early life environment, including social exposure to other children, influences the risk of asthma and allergy in a genotype-dependent manner (Hoffjan *et al.* 2005; Ober & Thompson 2005); indeed, *CCL5* attracts and stimulates histamine release in basophils, an important component of the pro-inflammatory allergic response (Laing & Secombes 2004).

Future work on the relationship between social environment, genotype and gene expression should aim at gaining a better understanding of the mechanisms that connect these factors. For example, social environment may be particularly important in shaping the expression of genes regulated by glucocorticoids (Cole *et al.* 2007; Miller *et al.* 2009). Dominance rank and glucocorticoid patterns are known to be linked in the Amboseli baboons: subordinate animals tend to exhibit higher rates of hypercortisolism than high-ranking individuals (Sapolsky *et al.* 1997), and a mother's dominance rank predicts glucocorticoid levels in her male offspring (Onyango *et al.* 2008). Further work on social environment-related GEIs should examine whether glucocorticoid circulation and signalling represent the physiological mechanism underlying these effects. At the same time, complementary work could explore whether epigenetic modification of the genome in response to early life environment accounts for long-term environmental effects, and whether GEIs represent cases in which environmentally responsive epigenetic modifications also vary depending on genetic variation.

ASGE measurements and genetic studies of natural populations

Our results help illustrate how ASGE measurements can be used as a tool for studying the genetic architecture of gene expression in natural populations. They help address one of the major challenges of conducting genetic investigations of wild populations—the inability to standardize the genetic and environmental background in which allelic variation operates—and present an opportunity to integrate environmental data on the same systems to investigate GEIs. Additionally, because of their focus on regulatory variation, field-based measurements of ASGE can be readily combined with laboratory-based tests for functional regulatory variation, should interesting candidates for further analysis be identified (i.e. transfection assays in cell culture or quantitative assays of transcription factor binding: Kurreeman *et al.* 2004; Tao *et al.* 2006; Zhu *et al.* 2006; Tung *et al.* 2009; Babbitt *et al.* 2010). These approaches can be used to test whether observed associations between ASGE and specific *cis*-regulatory variant(s) are causal or due to linkage disequilibrium with a nearby SNP.

What, then, will ASGE approaches be able to tell us about the evolutionary genetics of natural populations? One clear application is toward understanding the frequency, magnitude and environmental context-dependence of functional *cis*-regulatory variation in natural populations. Because such variation provides the raw material necessary for evolutionary change,

such survey-type approaches can yield important insight into which genes, gene sets or regions of the genome influence phenotypic variation that may be visible to natural selection. Additionally, understanding how environmental changes influence ASGE may help provide a partial mechanistic explanation for why the heritability of selectively relevant traits is sometimes observed to change under different environmental conditions (Qvarnstrom 1999; Wilson *et al.* 2006). Indeed, studies in laboratory model systems demonstrate that GEIs on gene expression can be pervasive, suggesting that the natural environmental variation experienced by animals in the wild plays an important role in modulating the genotype-gene expression relationship. As high throughput methodologies for measuring ASGE become more common (Degner *et al.* 2009; Fontanillas *et al.* 2010; Heap *et al.* 2010; Pastinen 2010), addressing these questions with functional data will become increasingly feasible.

In some cases, *cis*-regulatory genetic variation may also have a detectable effect on downstream, organism-level phenotypes, including fitness-related traits that can be measured in field studies. Given an a priori candidate locus, allelic imbalance assays have proved valuable in testing whether a *cis*-regulatory mechanism explains phenotypic variation within natural populations (Linnen *et al.* 2009; Tung *et al.* 2009) or between species (Wittkopp *et al.* 2009). Thus, when external evidence from quantitative trait mapping or association studies implicates one or a few specific loci, ASGE measurements can serve as an easy way to test for functional regulatory differences that might account for such a statistical link. It is somewhat less clear what role untargeted and/or genome-wide measurements of ASGE will play in understanding the genotype-phenotype relationship. Genes that exhibit allelic imbalance appear to be fairly common in the genome (Pastinen 2010), and ASGE data alone do not indicate the genes for which ASGE 'matters,' in the sense that expression imbalance contributes to downstream traits, vs. the genes for which ASGE simply reflects noise in the gene expression profile. This problem extends to gene expression and gene expression mapping studies more generally. A major challenge for this area therefore lies in understanding how the information embedded in ASGE and other expression data can be used in conjunction with other sources of data to dissect complex traits.

Conclusions

Our results demonstrate the potential of ASGE-based approaches for expanding the scope of gene expression studies from model organisms and cell culture, from which the majority of our current understanding arises, to ecologically well-characterized populations in the field. Such work has the potential to aid in understanding how genetic and environmental effects interact under field conditions, and to help in untangling the genetic basis of organism-level phenotypes of ecological or adaptive importance. For instance, our findings not only demonstrate that functional *cis*-regulatory variation is common in a natural population of nonhuman primates, but also suggest that multiple variants account for population-wide patterns of ASGE and that the strength of these effects is contingent on environmental variation. In particular, an early life environment, maternal dominance rank, appears to exert a long-term effect on the expression of a gene that plays an important role in mediating the inflammatory response. These results suggest that incorporating environmental effects of known importance in genetic studies of field populations will be critical to understanding the architecture of trait variation in the wild.

Acknowledgements

This work was supported by the National Science Foundation (BCS-0725502 to J.T. and S.C.A., BCS-0323553 to S.C.A, BCS-0323596 to J.A., BCS-0846286 to S.C.A. and G.A.W., and BCS-846532 to J.A.) and Duke University. We thank the Office of the President, Republic of Kenya, the Kenya Wildlife Service, its Amboseli staff and Wardens, the Institute of Primate Research, the National Museums of Kenya, and the members of the Amboseli-Longido pastoralist communities. Particular thanks go the Amboseli fieldworkers who contributed to genetic and environmental sampling, especially Raphael Mututua, Serah Sayialel, and Kinyua Warutere. We thank Tim Wango for invaluable support during sample collection efforts, Lacey Roerish and Niki Learn for database expertise and support, and the research staff and administration of the Institute of Primate Research, especially Tom Kariuki. Additionally, we are grateful to Sayan Mukherjee for help with the data analysis, Lisa Bukovnik for access to the pyrosequencer, and four anonymous referees for helpful comments on the manuscript. Sequence data have been deposited in GenBank under accession numbers HQ537531–HQ537665; sequence data analysed in this manuscript are also contained within previously submitted population data sets with accession numbers FJ952954–FJ953291, FJ953202–FJ953385, FJ953897–FJ954056, and FJ955224–FJ955397.

References

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) *Bmp4* and morphological variation of beaks in Darwin's finches. *Science*, **305**, 1462–1465.

Abzhanov A, Kuo WP, Hartmann C *et al.* (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**, 563–567.

Alberts SC, Altmann J (1995) Preparation and activation: determinants of age at reproductive maturity in male baboons. *Behavioral Ecology and Sociobiology*, **36**, 397–406.

Alberts SC, Altmann J (2001) Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *American Journal of Primatology*, **53**, 139–154.

Alberts SC, Buchan JC, Altmann J (2006) Sexual selection in wild baboons: from mating opportunities to paternity success. *Animal Behaviour*, **72**, 1177–1196.

Altmann SA (1991) Diets of yearling female primates (*Papio cynocephalus*) predict lifetime fitness. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 420–423.

Altmann J, Alberts SC (2003) Intraspecific variability in fertility and offspring survival in a nonhuman primate: behavioral control of ecological and social sources. In: *Offspring: The Biodemography of Fertility and Family Behavior* (eds Wachter K, Bulatao R), pp. 140–169. National Academy Press, Washington, D.C.

Altmann J, Alberts SC (2005) Growth rates in a wild primate population: ecological influences and maternal effects. *Behavioral Ecology and Sociobiology*, **57**, 490–501.

Altmann S, Altmann J (1970) *Baboon Ecology*. S. Karger, Basel.

Altmann J, Alberts SC, Haines SA *et al.* (1996) Behavior predicts genetic structure in a wild primate group. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 5797–5801.

Babbitt CC, Silverman JS, Haygood R *et al.* (2010) Multiple functional variants in *cis* modulate *PDYN* expression. *Molecular Biology and Evolution*, **27**, 465–479.

Bamshad MJ, Mummidi S, Gonzalez E *et al.* (2002) A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 10539–10544.

Barker DJP (2002) Fetal programming of coronary heart disease. *Trends in Endocrinology and Metabolism*, **13**, 364–368.

Barker DJP, Eriksson JG, Forsen T, Osmond C (2002) Fetal origins of adult disease: strength of effects and biological basis. *International Journal of Epidemiology*, **31**, 1235–1239.

Beehner JC, Nguyen N, Wango EO, Alberts SC, Altmann J (2006a) The endocrinology of pregnancy and fetal loss in wild baboons. *Hormones and Behavior*, **49**, 688–699.

Beehner JC, Onderdonk D, Alberts SC, Altmann J (2006b) The ecology of conception and pregnancy failure in wild baboons. *Behavioral Ecology*, **17**, 741–750.

Buchan JC, Alberts SC, Silk JB, Altmann J (2003) True paternal care in a multi-male primate society. *Nature*, **425**, 179–181.

Campbell CD, Kirby A, Nemesh J, Daly MJ, Hirschhorn JN (2008) A survey of allelic imbalance in F1 mice. *Genome Research*, **18**, 555–563.

Charpentier MJE, Tung J, Altmann J, Alberts SC (2008a) Age at maturity in wild baboons: genetic, environmental, and demographic influences. *Molecular Ecology*, **17**, 2026–2040.

Charpentier MJE, Van Horn RC, Altmann J, Alberts SC (2008b) Paternal effects on offspring fitness in a multi-male primate society. *Proceedings of the National Academy of Sciences*, **105**, 1988–1992.

Cheung VG, Bruzel A, Burdick JT *et al.* (2008) Monozygotic twins reveal germline contribution to allelic expression differences. *American Journal of Human Genetics*, **82**, 1357–1360.

Clutton-Brock TH (1989) *Red Deer in the Highlands*. Oxford University Press, Oxford, England.

Clutton-Brock TH, Pemberton J (2004) *Soay Sheep: Dynamics and Selection in an Island Population*. Cambridge University Press, Cambridge.

Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT (2007) Social regulation of gene expression in human leukocytes. *Genome Biology*, **8**, R189.

David P, Bjorksten T, Fowler K, Pomiankowski A (2000) Condition-dependent signalling of genetic variation in stalk-eyes flies. *Nature*, **406**, 186–188.

Degner JF, Marioni JC, Pai AA *et al.* (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, **25**, 3207–3212.

Fontanillas P, Landry CR, Wittkopp PJ *et al.* (2010) Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Molecular Ecology*, **19**, 212–227.

Gesquiere LR, Wango EO, Alberts SC, Altmann J (2007) Mechanisms of sexual selection: sexual swellings and estrogen concentrations as fertility indicators and cues for male consort decisions in wild baboons. *Hormones and Behavior*, **51**, 114–125.

Godfrey KM, Barker DJP (2000) Fetal nutrition and adult disease. *American Journal of Clinical Nutrition*, **71**, 1344S–1352S.

Grant PR (1986) *Ecology and Evolution of Darwin's Finches*. Princeton University Press, Princeton, NJ.

Gruber JD, Long AD (2009) *Cis*-regulatory variation is typically polyallelic in *Drosophila*. *Genetics*, **181**, 661–670.

Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *American Journal of Human Genetics*, **66**, 1669–1679.

Hamblin MT, Thompson EE, di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. *American Journal of Human Genetics*, **70**, 369–383.

Heap GA, Yang JH, Downes K *et al.* (2010) Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Human Molecular Genetics*, **19**, 122–134.

Hoffjan S, Nicolae D, Ostrovnaya I *et al.* (2005) Gene–environment interaction effects on the development of immune responses in the 1st year of life. *American Journal of Human Genetics*, **76**, 696–704.

Hughes AL, Packer B, Welch R, Chanock SJ, Yeager M (2005) High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. *Immunogenetics*, **57**, 821–827.

Johnson SE (2003) Life history and the competitive environment: trajectories of growth, maturation, and reproductive output among Chacma baboons. *American Journal of Physical Anthropology*, **120**, 83–98.

von Korff M, Radovic S, Choumane W *et al.* (2009) Asymmetric allele-specific expression in relation to developmental variation and drought stress in barley hybrids. *The Plant Journal*, **59**, 14–26.

Kruuk LE, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings. Biological Sciences/The Royal Society*, **275**, 593–596.

Kurreeman FA, Schonkeren JJ, Heijmans BT, Toes RE, Huizinga TW (2004) Transcription of the *IL10* gene reveals

allele-specific regulation at the mRNA level. *Human Molecular Genetics*, **13**, 1755–1762.
Laing KJ, Secombes CJ (2004) Chemokines. *Developmental & Comparative Immunology*, **28**, 443–460.
Lindstrom J (1999) Early development and fitness in birds and mammals. *Trends in Ecology & Evolution*, **14**, 343–348.
Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science*, **325**, 1095–1098.
McGregor AP, Orgogozo V, Delon I *et al.* (2007) Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature*, **448**, 587–590.
McKusick-Nathans Institute of Genetic Medicine JHU, National Center for Biotechnology Information NLoM (2010) Online Mendelian Inheritance in Man (OMIM).
Meaney MJ, Szyf M (2005) Maternal care as a model for experience-dependent chromatin plasticity? *Trends in Neurosciences*, **28**, 456–463.
de Meaux J, Goebel U, Pop A, Mitchell-Olds T (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of *Arabidiopsis thaliana* that is compatible with neutral evolution. *The Plant Cell*, **17**, 676–690.
de Meaux J, Pop A, Mitchell-Olds T (2006) *Cis*-regulatory evolution of chalcone-synthase expression in the genus *Arabidopsis*. *Genetics*, **174**, 2181–2202.
Milani L, Gupta M, Andersen M *et al.* (2007) Allelic imbalance in gene expression as a guide to *cis*-acting regulatory single nucleotide polymorphisms in cancer cells. *Nucleic Acids Research*, **35**, e34.
Miller GE, Chen E, Fok AK *et al.* (2009) Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 14716–14721.
Morley M, Molony CM, Weber TM *et al.* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
Ober C, Thompson EE (2005) Rethinking genetic models of asthma: the role of environmental modifiers. *Current Opinion in Immunology*, **17**, 670–678.
Onyango PO, Gesquiere LR, Wango EO, Alberts SC, Altmann J (2008) Persistence of maternal effects in baboons: mother's dominance rank at son's conception predicts stress hormone levels in subadult males. *Hormones and Behavior*, **54**, 319–324.
Pant PVK, Tao H, Beilharz EJ *et al.* (2006) Analysis of allelic differential expression in human white blood cells. *Genome Research*, **16**, 331–339.
Pastinen T (2010) Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews. Genetics*, **11**, 533–538.
Pastinen T, Hudson TJ (2004) *Cis*-acting regulatory variation in the human genome. *Science*, **306**, 647–650.
Qvarnstrom A (1999) Genotype-by-environment interactions in the determination of the size of a secondary sexual character in the collared flycatcher (*Ficedula albicollis*). *Evolution*, **53**, 1564–1572.
R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Ravelli GP, Stein ZA, Susser MW (1976) Obesity in young men after famine exposure in utero and early infancy. *New England Journal of Medicine*, **295**, 349–353.
Reifsnyder PC, Churchill G, Leiter EH (2000) Maternal environment and genotype interact to establish diabesity in mice. *Genome Research*, **10**, 1568–1578.
Samuels A, Altmann J (1986) Immigration of a *Papio anubis* male into a group of *Papio cynocephalus* baboons and evidence for an anubis-cynocephalus hybrid zone in Amboseli, Kenya. *International Journal of Primatology*, **7**, 131–138.
Sapolsky RM, Alberts SC, Altman J (1997) Hypercortisolism associated with social subordinance or social isolation among wild baboons. *Archives of General Psychiatry*, **54**, 1137–1143.
Schaart JG, Mehli L, Schouten HJ (2005) Quantification of allele-specific expression of a gene encoding strawberry polygalacturonase-inhibiting protein (PGIP) using Pyrosequencing. *Plant Journal*, **41**, 493–500.
Schalkwyk LC, Meaburn EL, Smith R *et al.* (2010) Allelic skewing of DNA methylation is widespread across the genome. *American Journal of Human Genetics*, **86**, 196–212.
Serre D, Gurd S, Ge B *et al.* (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic *cis*-acting mechanisms regulating gene expression. *PLoS Genetics*, **4**, e1000006.
Silk JB, Alberts SC, Altmann J (2003) Social bonds of female baboons enhance infant survival. *Science*, **302**, 1231–1234.
Silk JB, Beehner JC, Bergman TJ *et al.* (2009) The benefits of social capital: close social bonds among female baboons enhance offspring survival. *Proceedings. Biological Sciences/The Royal Society*, **276**, 3099–3104.
St Clair D, Xu MQ, Wang P *et al.* (2005) Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. *Journal of the American Medical Association*, **294**, 557–562.
Suzuki Y, Hattori S, Mashimo Y *et al.* (2009) *CD14* and *ILR4R* gene polymorphisms modify the effect of day care attendance on serum IgE levels. *Journal of Allergy and Clinical Immunology*, **123**, 1408–1411.
Tao H, Cox DR, Frazer KA (2006) Allele-specific *KRT1* expression is a complex trait. *PLoS Genetics*, **2**, 848–858.
Tung J, Charpentier MJE, Garfield DA, Altmann J, Alberts SC (2008) Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Molecular Ecology*, **17**, 1998–2011.
Tung J, Primus A, Bouley AJ *et al.* (2009) Evolution of a malaria resistance gene in wild primates. *Nature*, **460**, 388–392.
Tung J, Alberts SC, Wray GA (2010) Evolutionary genetics in wild primates: combining genetic approaches with field studies of natural populations. *Trends in Genetics*, **26**, 353–362.
Veyrieras JB, Kudaravalli S, Kim SY *et al.* (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genetics*, **4**, e1000214.
Walters JR (1987) Transition to adulthood. In: *Primate Societies* (eds Smuts BB, Cheney DL, Seyfarth R, Wrangham RW, Struhsaker TT), pp. 358–369. University of Chicago Press, Chicago, IL.

Weaver ICG, Cervoni N, Champagne FA *et al.* (2004) Epigenetic programming by maternal behavior. *Nature Neuroscience*, **7**, 847–854.

Wilson AJ, Pemberton JM, Pilkington JG *et al.* (2006) Environmental coupling of selection and heritability limits evolution. *PLoS Biology*, **4**, e216.

Wittkopp P, Haerum B, Clark A (2004) Evolutionary changes in cis and trans gene regulation. *Nature*, **430**, 85–88.

Wittkopp P, Haerum B, Clark AG (2008) Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genetics*, **40**, 346–350.

Wittkopp PJ, Stewart EE, Arnold LL *et al.* (2009) Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science*, **326**, 540–544.

Wooding S, Bufe B, Grassi C *et al.* (2006) Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature*, **440**, 930–934.

Wray GA, Hahn MW, Abouheif E *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, **20**, 1377–1419.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW (2002) Allelic variation in human gene expression. *Science*, **297**, 1143.

Zhang X, Borevitz JO (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics*, **182**, 943–954.

Zhu CY, Odeberg J, Hamsten A, Eriksson P (2006) Allele-specific *MMP-3* transcription under in vivo conditions. *Biochemical and Biophysical Research Communications*, **348**, 1150–1156.

J.T. is interested in the evolutionary genetics of natural populations, particularly the interplay between ecology, behaviour, and genetics. M.A. is a research scientist at the Institute of Primate Research in Kenya, and is interested in the relationship between behaviour, genetics, and disease in nonhuman primates. S.M. also works at IPR, where he is a research technician interested in molecular biology and genetics. G.W. is an evolutionary biologist who is broadly interested in the connection between genetic variation and its evolutionary consequences. S.A. and J.A. are co-directors of the Amboseli Baboon Research Project. Their research program seeks to identify demographic, environmental, endocrine and genetic correlates of and sources of variance in fitness components in a natural population of mammals.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Supplementary materials, methods and references.

Table S1 Correlations between *CCL5* measurements obtained under different sample storage conditions.

Tables S2–S5 Year of sampling (e.g. batch) effects on ASGE measurements for genes that exhibit common allelic imbalance in the Amboseli population.

Tables S6–S10 Correlations (R^2) between heterozygosity/homozygosity at different sites in the sequenced *cis*-regulatory regions and consensus sequence for each locus for individuals in the ASGE sample set, with variable sites in bold (abbreviated using IACUC ambiguity codes).

Fig. S1 ASGE measurements are consistent across sample handling treatments for (a) *CCL5* ($n = 7$ individuals), a gene that exhibits common, large-scale ASGE; (b) *CXCR4* ($n = 6$ individuals), a gene that exhibits significant ASGE at smaller scales; and (c) *TAP2* ($n = 2$ individuals), for which we detected no significant common ASGE.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.