REforge Associates Transcription Factor Binding Site Divergence in Regulatory Elements with Phenotypic Differences between Species

Björn E. Langer,[1,2,3] Juliana G. Roscito,[1,2,3] and Michael Hiller*,[1,2,3]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[2]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[3]Center for Systems Biology, Dresden, Germany

***Corresponding author:** E-mail: hiller@mpi-cbg.de.
Associate editor: Naruya Saitou

Abstract

Elucidating the genomic determinants of morphological differences between species is key to understanding how morphological diversity evolved. While differences in *cis*-regulatory elements are an important genetic source for morphological evolution, it remains challenging to identify regulatory elements involved in phenotypic differences. Here, we present Regulatory Element forward genomics (REforge), a computational approach that detects associations between transcription factor binding site divergence in putative regulatory elements and phenotypic differences between species. By simulating regulatory element evolution in silico, we show that this approach has substantial power to detect such associations. To validate REforge on real data, we used known binding motifs for eye-related transcription factors and identified significant binding site divergence in vision-impaired subterranean mammals in 1% of all conserved noncoding elements. We show that these genomic regions are significantly enriched in regulatory elements that are specifically active in mouse eye tissues, and that several of them are located near genes, which are required for eye development and photoreceptor function and are implicated in human eye disorders. Thus, our genome-wide screen detects widespread divergence of eye-regulatory elements and highlights regulatory regions that likely contributed to eye degeneration in subterranean mammals. REforge has broad applicability to detect regulatory elements that could be involved in many other phenotypes, which will help to reveal the genomic basis of morphological diversity.

Key words: transcription factor binding site, regulatory element, comparative genomics.

Introduction

Due to advances in sequencing technology, the number of sequenced genomes is rapidly increasing. This wealth of genomic data provides unprecedented opportunities to identify the genomic changes that underlie particular phenotypic differences between the sequenced species, which is an important challenge in genomics and evolutionary biology. In particular, numerous sequenced genomes make it possible to extend association-based approaches, which were instrumental in discovering intraspecies genotype–phenotype associations (MacArthur et al. 2017), to cross-species comparisons.

One such approach to link genomic differences between species to phenotypes is the Forward Genomics framework (Hiller et al. 2012; Prudent et al. 2016). This approach relies on two concepts. First, genomic information that is necessary for a particular phenotype should be largely conserved in species where this phenotype is present and under selection. In contrast, this genomic information should evolve under relaxed selection or neutrally in species that have lost this phenotype. Second, the repeated loss of a phenotype in independent lineages is expected to result in a specific signature, where

the genomic regions uniquely necessary for this phenotype are preferentially diverged in those species that lack the phenotype. The focus on phenotypes that have repeatedly changed is key to obtaining specificity in a genome-wide search (Hiller et al. 2012). By systematically investigating all genomic regions that are overall conserved among the considered species, Forward Genomics screens for a match between the repeated sequence divergence signature and the repeated phenotypic change. To measure sequence divergence on a per-species basis, the original Forward Genomics method requires a phylogenetic tree and a multiple alignment of an individual genomic region, and uses Maximum Likelihood to reconstruct the sequence of the common ancestor. Then, it calculates sequence divergence as the percent of nucleotide changes that happened between the ancestral sequence and the sequence of an extant species, considering both substitutions and insertions/deletions (Hiller et al. 2012). Since Forward Genomics only requires genomic sequence data and a phylogeny, it can be applied to many phenotypes that repeatedly changed in species with sequenced genomes.

The reliance on "plain nucleotide divergence" allows the standard Forward Genomics method to run an unbiased

Open Access

Article

genome-wide screen that considers all conserved genomic regions, regardless of whether they overlap coding exons, regulatory elements or other classes of functional elements. The same holds for a recently developed approach, which compares divergence in genomic regions by estimating element-specific phylogenetic branch lengths that are proportional to the number of occurred substitutions (Partha et al. 2017), and also for the Reverse Genomics method, which uses percent sequence identity and percent sequence match (number of nucleotide matches divided by alignment length and number of reference bases, respectively) to classify conserved noncoding regions as "conserved" and "lost" in different species (Marcovitz et al. 2016). The drawback of this generality is that plain sequence divergence, regardless of whether insertions/deletions are considered or not, ignores the underlying grammar that determines the function for specific types of elements. For cis-regulatory elements (CREs), the "words" are binding sites for transcription factors (TFs) that bind to the element and control expression of the target gene. It is therefore possible that a CRE largely preserves its regulatory activity if changes predominantly occur outside of transcription factor binding sites (TFBS) or maintain the binding preference of a TF (binding motif), as illustrated in figure 1A. In addition, TFBS turnover, a frequent process where mutations create a new binding site followed by the loss of the ancestral binding site (Huang et al. 2007; Otto et al. 2009; Villar et al. 2014), results in increased sequence divergence that may not affect regulatory activity. On the other hand, minor sequence changes that preferentially affect key TFBSs may substantially affect regulatory activity, but would not stand out in a genome-wide search for sequence divergence (fig. 1B). Consequently, sequence conservation and divergence is not an ideal indicator of functional conservation and divergence of regulatory elements. Therefore, the unbiased applicability of standard Forward Genomics or similar approaches comes at the cost of a reduced sensitivity and specificity in discovering divergence in regulatory elements that likely affects regulatory activity in species with the altered trait.

Functional differences in CREs are an important evolutionary source for phenotypic change. In particular, it is thought that morphology largely evolves by changes in CREs that affect the spatio-temporal expression of key developmental genes (Carroll 2005, 2008; Wray 2007). A number of studies have linked CRE divergence with morphological traits such as different pigmentation patterns in Drosophilids (Jeong et al. 2008), loss of limbs in snakes (Kvon et al. 2016; Leal and Cohn 2016), loss of pelvic spines and armor plate reductions in freshwater stickleback (Chan et al. 2010; O'Brown et al. 2015), wing development in bats (Cretekos et al. 2008; Booker et al. 2016), and human traits such as the absence of pelvic spines and toe size reductions (McLean et al. 2011; Indjeian et al. 2016). Thus, in order to use comparative genomics to associate CRE divergence with phenotypes, it would be desirable to have an approach that specifically detects sequence changes in CREs that likely affect regulatory activity and associates such changes with phenotypes.

Here, we present Regulatory Element forward genomics (REforge), a new method to predict CREs that are involved in a phenotypic change by measuring differences in binding sites of TFs that are relevant for this phenotype. As demonstrated on large CRE sets obtained by simulating regulatory element evolution and trait loss, REforge has greatly improved power to detect divergence in CREs that are associated with a phenotypic difference. To validate the method on real data, we applied REforge to detect CREs that exhibit preferential divergence of eye-related TFBSs in vision-impaired subterranean mammals. We show that the genomic regions uncovered in this genome-wide screen are significantly enriched in eye-specific regulatory elements. Furthermore, several of these regions are located near genes with key roles in eye development and function as well as genes implicated in human eye disorders, suggesting that TFBS divergence in these eye-regulatory elements could be involved in eye degeneration in subterranean species. In general, REforge has broad applicability to other traits to link binding site divergence in CREs with phenotypic differences of the sequenced species.

Results

REforge Method

The main rationale behind REforge is to replace sequence divergence by a divergence measurement that captures differences in the important building blocks of CREs (TFBS), and thus better predicts functional differences in regulatory activity. Like standard Forward Genomics, REforge assumes that the CREs that are important for a trait should maintain binding sites for relevant TFs in species that possess this trait, although not necessarily at conserved positions. In species in which the trait is absent, sequence divergence that is a sign of functional divergence should destroy or weaken many important TFBSs, in contrast to sequence divergence that largely preserves relevant TFBSs and thus likely regulatory activity.

REforge requires as input 1) a set of putative CREs and their orthologous aligning sequences in a set of species, 2) a phylogenetic tree, 3) a list of all species in which the phenotype is lost, and 4) a set of motifs for TFs that are likely relevant for the phenotype. Such TFs can be obtained by intersecting TFs with Gene Ontology terms that relate to the phenotype. Alternatively, TFs which are highly expressed in the respective tissues or which affect the phenotype in a mouse knockout represent good candidates. Motifs that reflect the binding preferences of these TFs can be obtained from TRANSFAC (Matys et al. 2006), JASPAR (Mathelier et al. 2014), UniPROBE (Hume et al. 2015), or other sources.

The first step in quantifying conservation or divergence of TFBSs in a given DNA sequence is to compute a score that reflects the collective binding affinity of the set of relevant TFs (fig. 1C). To this end, we made use of Stubb (Sinha et al. 2003, 2006), a Hidden Markov model-based method that takes a set of TF motifs as input. Stubb has already been successfully used to find correlations between TFBSs and phenotypes (Kapheim et al. 2015) and has the following advantages. First, Stubb avoids overcounting the contribution of overlapping binding sites that can only be occupied by one TF at the time. Second, Stubb considers both strong and weak binding
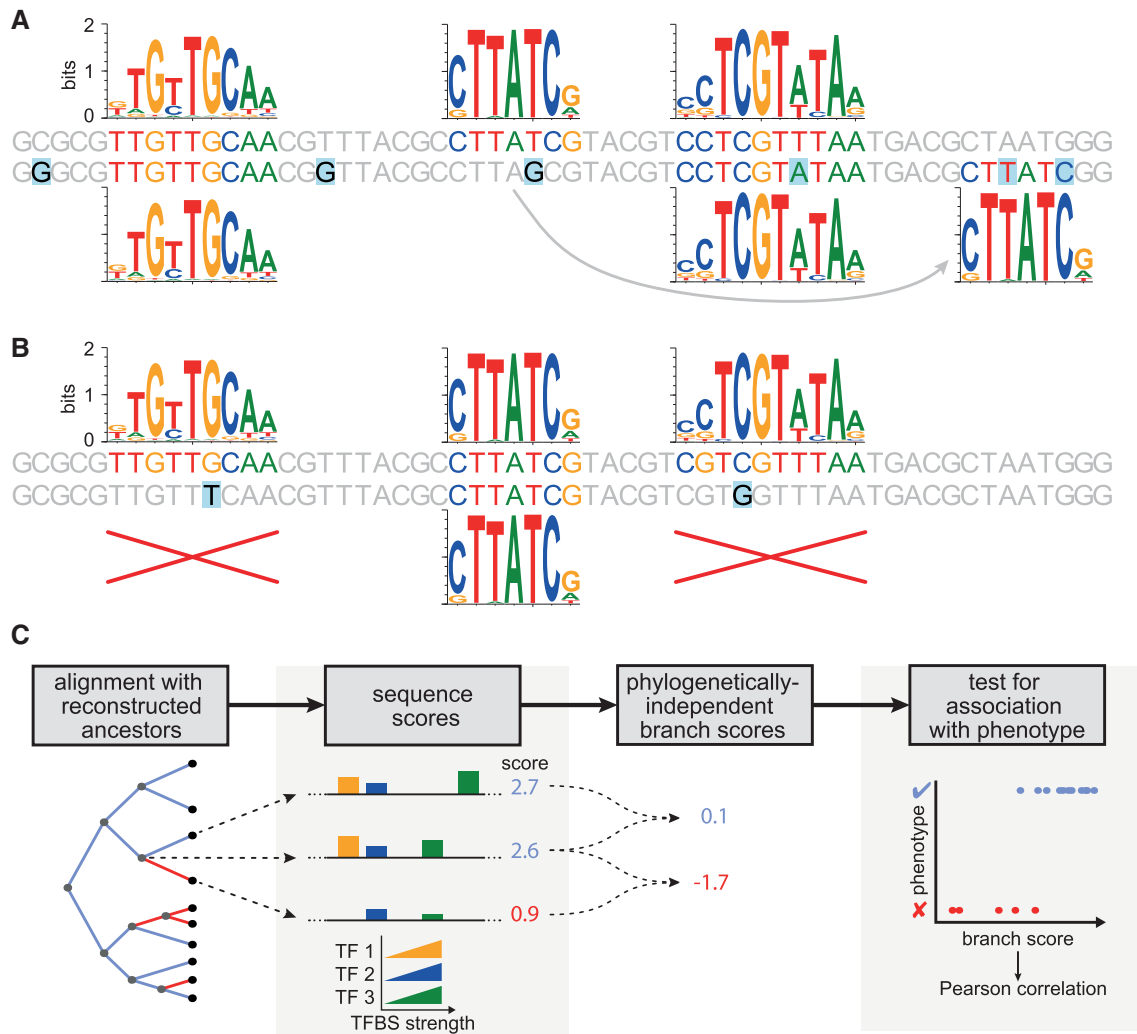
FIG. 1. REforge incorporates knowledge about transcription factor binding sites to associate divergence in *cis*-regulatory elements with phenotypic differences. (A) Illustration of multiple sequence changes (blue background) that largely preserve the TFBS ensemble of two sequences. The mutations are either located outside of binding sites, preserve the motif or result in TFBS turnover (gray arrow). (B) Illustration of just two sequence changes that destroy TFBS and thus are more likely to result in functional divergence. (C) Overview of REforge. For each node in the phylogenetic tree, REforge computes sequence scores that reflect the collective binding affinity of the given TF set. Branch scores, which are computed as the difference between the sequence scores, reflect TFBS changes. Finally, REforge tests if the branch scores of trait-loss branches (red) are lower than the branch scores of trait-preserving branches (blue).

sites without relying on fixed thresholds, which is important since weak TFBSs contribute to the precision of expression patterns driven by developmental enhancers (Farley et al. 2015). Third, since Stubb does not explicitly consider the position of a motif, there are no constraints on the arrangement of TFBS. Consequently, Stubb scores are not sensitive to TFBS turnover.

We found that random sequences, which contain TFBSs only by chance, typically have nonzero Stubb scores (supplementary fig. 1, Supplementary Material online). To ensure that scores for such sequences are on an average zero, we normalized the Stubb score by subtracting from it the average score of ten sequences that were obtained by shuffling the nucleotides of the given sequence (supplementary fig. 2, Supplementary Material online). The resulting score is called the "sequence score" (fig. 1C).

Next, to quantify the extent to which the estimated binding affinity of the TF set changed during evolution, we adopted the Forward Genomics branch method (Prudent et al. 2016). This method requires as input a multiple alignment of a genomic region with reconstructed ancestral sequences for every internal node in the given phylogeny. We then consider the sequence at the start (an ancestral sequence) and end of a branch (either the sequence of an ancestor or an extant species) and compute the difference between the two sequence scores (fig. 1C), which is called the "branch score." If TFBSs are largely preserved or even increased along this branch, the branch score will be ~0 or positive. In contrast, if mutations destroyed or weakened TFBS, the overall binding affinity will be lower at the branch end and the branch score will be negative.

Finally, to test for an association between binding site divergence in a CRE and a given phenotypic difference, we test if the branches leading to nodes or extant species that have lost the trait have significantly lower branch scores compared with the remaining branches (fig. 1C). Since every branch in a phylogeny represents independent evolution, the branch scores are phylogenetically independent and no further correction for phylogenetic dependencies between species is necessary. We compared different statistical tests (Pearson and Spearman correlation, Wilcoxon rank-sum, and t-test), which showed that the significance of a positive Pearson correlation coefficient, computed between the branch scores and the binary branch classification (trait-loss vs. trait-preserving branches), performed best (supplementary fig. 3, Supplementary Material online).

REforge Substantially Outperforms Standard Forward Genomics

To test the performance of REforge and to compare it with standard Forward Genomics, we used PEBCRES to simulate CRE evolution (Duque et al. 2014) and generated a large data set of CREs where the ground truth is known. In this simulation, selection acts on a predefined regulatory activity pattern of a CRE, which is inferred from binding sites of five TFs in the sequence (supplementary fig. 4A, Supplementary Material online). Thus, during in silico evolution, the fitness of each CRE is maintained if sequence changes do not affect binding sites or if equivalent TFBS arise (turnover). We considered a 20-species phylogeny that contained three independent trait-loss species (fig. 2A) and generated two sets of CREs. First, we computationally evolved 10,000 CREs (background) that are under selection to preserve the regulatory activity pattern along every branch. Since these background CREs are not associated with the lost phenotype and any preferential binding site divergence along the branches leading to trait-loss species would be due to random chance alone, these CREs were counted as negatives. Second, we computationally evolved 200 CREs (foreground) that are associated with the lost trait. These CREs evolve under selection throughout the tree, except for the final part of three branches leading to the trait-loss species, where these CREs then evolved without selection to preserve regulatory activity (fig. 2A). These foreground CREs are associated with the loss of the phenotype and are counted as positives. Overall, the entire set of 10,200 CREs contains 1.96% positives.

We generated three sets of 200 foreground CREs that differ in the age of the trait loss by setting the length of the final part of the trait-loss branches to 0.03, 0.06, and 0.09 substitutions per site, respectively. We found that REforge is able to distinguish the 200 foreground CREs from the 10,000 background CREs with an area under the receiver operating characteristic curve (AUC) of 0.9344, 0.9952, and 0.9993 for the three trait loss ages (fig. 2B). In contrast, standard Forward Genomics achieves a lower AUC of 0.5420, 0.5859, and 0.6209 on the same data sets. Comparing the area under a precision-sensitivity curve (see following paragraph) corroborates this result (supplementary fig. 5, Supplementary Material online). This shows that REforge substantially outperforms standard

Forward Genomics in detecting simulated CREs that have diverged binding sites on the trait-loss branches.

REforge is Robust to Various Parameters

Next, we evaluated the robustness of REforge to various parameters that are relevant for application to real data. We expect that, in general, the vast majority of CREs are not involved in the loss of the trait and thus are negatives. For such strongly imbalanced data sets, recapitulated in our simulation by >98% negatives, the use of receiver operating characteristics, which compares sensitivity and specificity (proportion of correctly identified negative CREs; 1—false positive rate), is not appropriate (Saito and Rehmsmeier 2015). Since the main objective of REforge is to achieve a high precision (defined as the proportion of trait-involved positive CREs out of all identified CREs; also called positive predictive value), we assessed performance in the following by comparing sensitivity versus precision.

Since the age of loss will differ between traits and between trait-loss species, we first reconsidered the sensitivity versus precision performance of REforge on the three data sets that differ by the age of the trait-loss. Considering a high precision of 90%, REforge achieves a sensitivity of 29%, 79%, and 96%, respectively, for the three ages (fig. 2C). This shows that the older the trait loss, the easier it is to detect foreground CREs, which is expected as CREs that evolve neutrally for a longer time will show a higher binding site divergence. Nevertheless, REforge is able to identify a certain percentage of positive CREs at 90% precision also for trait losses that occurred relatively recently (fig. 2C).

Second, for real data applications the ancestral sequences are not known, as in our simulation, but have to be reconstructed. Furthermore, less than 20 species may be available. Therefore, we tested REforge considering a phylogeny of only eight trait-preserving and three trait-loss species (fig. 2A, asterisks) and reconstructed ancestral sequences using PRANK (Loytynoja and Goldman 2008). Compared with using a 20-species phylogeny and known ancestral sequences, the sensitivity at a precision of 90% slightly drops by 4–9% (fig. 2C). However, REforge still achieves a sensitivity of 20%, 75%, and 92% for the three trait-loss ages, showing that the method is robust toward ancestral reconstruction uncertainty on a phylogeny with less species.

Third, when REforge is applied to real data, it is likely that the input set of CRE candidates also contains CREs that are active in tissues, which are not relevant for the lost trait. This is especially the case when REforge is applied to a set of conserved noncoding regions (see below), where many regions will not overlap CREs active in the tissues of interest, or may not overlap CREs at all. So far, both the foreground and background set of CREs have activity in a tissue relevant for the lost trait and this activity was controlled by binding sites for the same five TFs (supplementary fig. 4A, Supplementary Material online). Therefore, we assessed the performance of REforge on a data set where 5,000 of the original background CREs were replaced by 5,000 new background CREs that are active in an unrelated tissue and where this activity is controlled by five other TFs (supplementary fig. 4B,
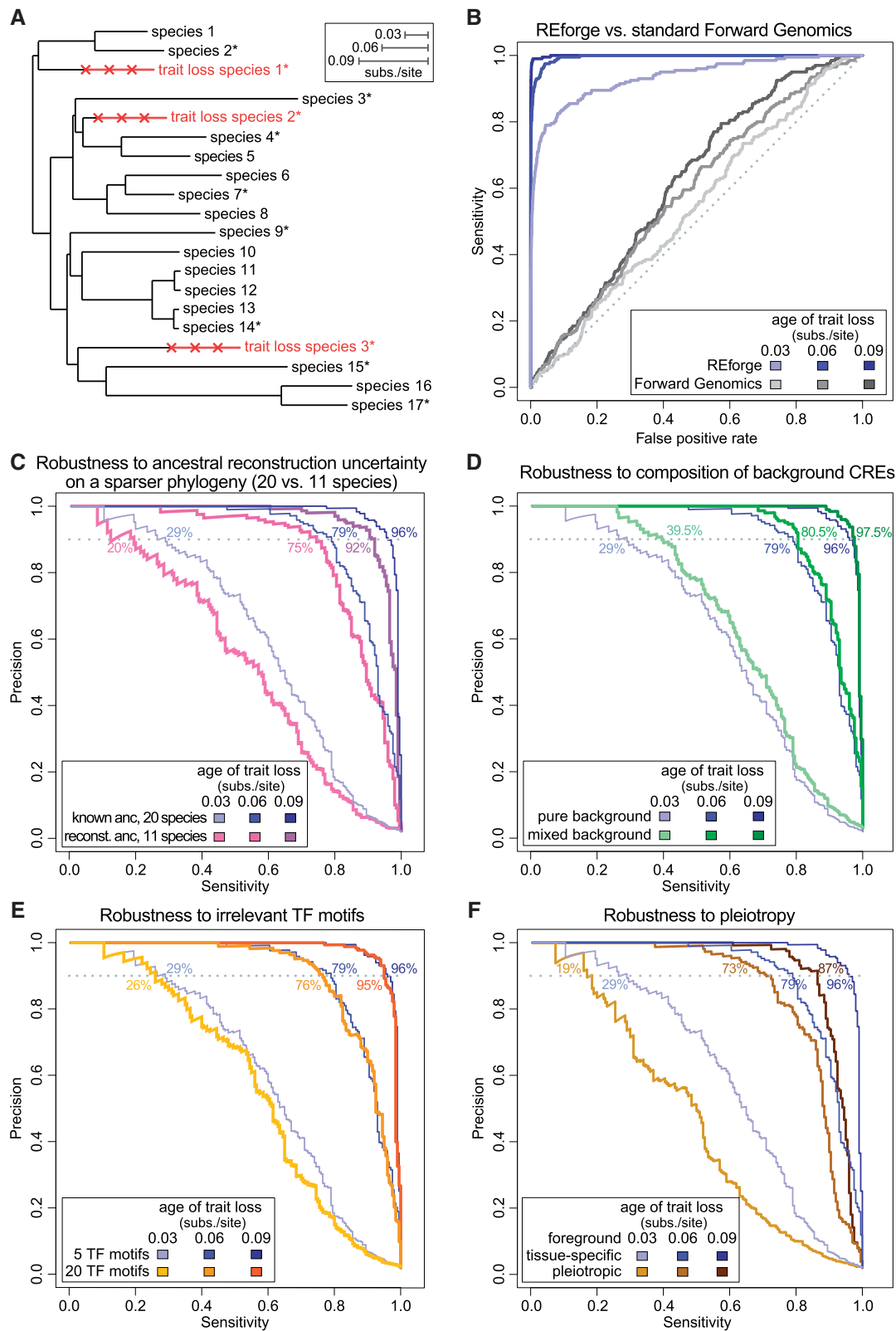
FIG. 2. REforge outperforms standard Forward Genomics and is robust to various simulation parameters. (*A*) Phylogenetic tree that was used to simulate the evolution of 200 foreground and 10,000 background CREs. The terminal parts of the three trait-loss branches (red) that correspond to the three trait-loss ages (0.03, 0.06, and 0.09 substitutions per site) are indicated by red crosses. An asterisk marks those species that were included in the 11-species phylogeny (panel *C*). (*B*) Receiver operating characteristic curves (sensitivity vs. false positive rate) compare REforge and standard Forward Genomics for three trait-loss ages. (*C*) Tests that compare a 20-species phylogeny and known ancestral sequences (shades of blue) and a sparser phylogeny of 11 species for which ancestral sequences were reconstructed. Blue curves serve as reference in panels (*C–F*) and show the same data as in panel (*B*) but in a sensitivity versus precision plot. (*D*) Tests that include background CREs that are not active in the tissues relevant

Supplementary Material online). Compared with the original background CRE set, REforge achieved on this mixed background set a similar sensitivity at 90% precision for the medium and old trait loss scenarios and a ~10% higher sensitivity for recent trait losses (fig. 2D), showing that the new background CREs, which have different TFBSs, are easier to recognize as negatives.

Fourth, if functional annotations such as Gene Ontology, expression or gene-knockout data is used to select promising TFs as input to REforge, it is likely that not all these TFs are actually relevant for the lost trait. Therefore, we tested the performance of REforge on input sets of TF motifs that included not only the five TFs controlling CRE activity in the relevant tissue but also 15 other TF motifs that are not important for the activity of these simulated CREs (supplementary table 1, Supplementary Material online). As shown in figure 2E and supplementary figure 6, Supplementary Material online, while the motifs of 15 unrelated TFs introduce noise, REforge's performance is largely similar and sensitivity only slightly decreases by 1–3% at a precision of 90%.

Fifth, CREs can be pleiotropic and control expression in different tissues, as shown by enhancers with shared activity in the developing limb and genital tissues (Infante et al. 2015). After trait loss, pleiotropic CREs are expected to lose selection for TFBSs that are only required for regulatory activity in tissues related to the lost trait. However, pleiotropic CREs should still be under selection to preserve TFBSs required for regulatory activity in the other tissues. Thus, while some TFBSs may get lost, the overall divergence is more limited in pleiotropic CREs. To explore how the performance of REforge is affected by pleiotropy, we simulated pleiotropic CREs that are active in two tissues, controlled by two different sets of five TFs each (supplementary fig. 4, Supplementary Material online). The first tissue is relevant for the lost trait, while the second tissue is not. After trait-loss, selection pressure on the pleiotropic foreground CREs changed to preserve only regulatory activity in the second tissue. Compared with 200 tissue-specific foreground CREs, applying REforge to 10,200 CREs that contain 200 pleiotropic foreground CREs decreases sensitivity by 6–10% at a precision of 90% for the three trait-loss ages (fig. 2F). This shows that while it is harder to detect binding site divergence in pleiotropic CREs, REforge is capable of identifying trait-involved pleiotropic CREs in a large set of CREs.

In summary, while the age of the trait loss has a major influence on the performance, several other factors such as number of species, ancestral reconstruction uncertainty, the presence of background CREs with binding sites for different TFs, the addition of unrelated TF motifs, or pleiotropy of the foreground CREs have a rather small influence. We conclude that at a high precision of 90% REforge is able to detect a

sizeable portion of simulated CREs with preferential binding site divergence along the branches leading to trait-loss species.

Genome-Wide Application of REforge to Uncover CREs Involved in Eye Degeneration

To test REforge on real data, we focused on eye degeneration as a trait loss for the following reasons. First, four independent subterranean mammals (blind mole-rat, naked mole-rat, star-nosed mole, and cape golden mole; fig. 3A) that possess small, rudimentary eyes have sequenced genomes (Nevo 1979; Sanyal et al. 1990; Catania 1999; Hetling et al. 2005; Nemec et al. 2008). Second, many TFs that are important for eye development are known. Third, genome-wide regulatory data sets such as enhancer marks or TF bound regions are available for several mouse eye tissues. These data sets can be used to validate that genomic regions detected by REforge overlap eye-related regulatory elements.

We applied REforge to a genome-wide set of 351,279 conserved noncoding elements (CNEs) obtained from a multiple genome alignment of the four aforementioned subterranean species, 22 other mammals, and lizard, chicken, and frog as outgroup species (fig. 3A). Since CNEs often overlap CREs (Woolfe et al. 2005; Pennacchio et al. 2006; Capra et al. 2013), these genomic regions are a good proxy for regulatory elements. To obtain input TF motifs, we compiled a set of 28 motifs for TFs (supplementary table 2, Supplementary Material online) that are known to be important for lens or retina development (Cvekl and Mitton 2010). Screening for preferential binding site divergence in the subterranean mammals, REforge identified 3,711 (1.06%) of the 351,279 CNEs at a false discovery rate of 10%. Supplementary table 3, Supplementary Material online, lists the coordinates and P values of these 3,711 CNEs, the distance to the transcription start site of the closest upstream and downstream genes, and functional annotations of these genes.

To explore if these 3,711 CNEs overlap CREs that are active in eye tissues in species with normal eyes, we compiled a list of publicly available eye-regulatory data sets and determined the overlap with these CNEs. To assess if the observed overlap is significantly higher than expected by chance, we used a Fisher's exact test and additionally sampled 10,000 random sets of 3,711 CNEs each from the entire set of 351,279 CNEs. These random sets were used to compute a Z-score that measures how many SDs the observed overlap is above or below the average of the 10,000 random sets.

We found a significantly higher overlap between the 3,711 REforge-identified CNEs and a number of eye-related regulatory data sets from mouse retina and lens tissues (fig. 3B and supplementary fig. 7, Supplementary Material online). These data sets include genomic regions bound by essential eye

FIG. 2. Continued
for the lost trait. Pure background refers to 10,000 background CREs whose activity is controlled by the same five TFs as the 200 foreground CREs. The pure background data sets are identical to the reference in panel (C). Mixed background refers to a different background set where the activity of 5,000 of 10,000 CREs is controlled by five different TFs. (E) Tests that include 15 unrelated TFs that do not influence the activity of foreground or background CREs. The 10,000 pure background set was used. (F) Tests on data sets that contain 200 pleiotropic foreground CREs having activity in two tissues, of which only one is relevant for the lost trait. The 10,000 pure background set was used.
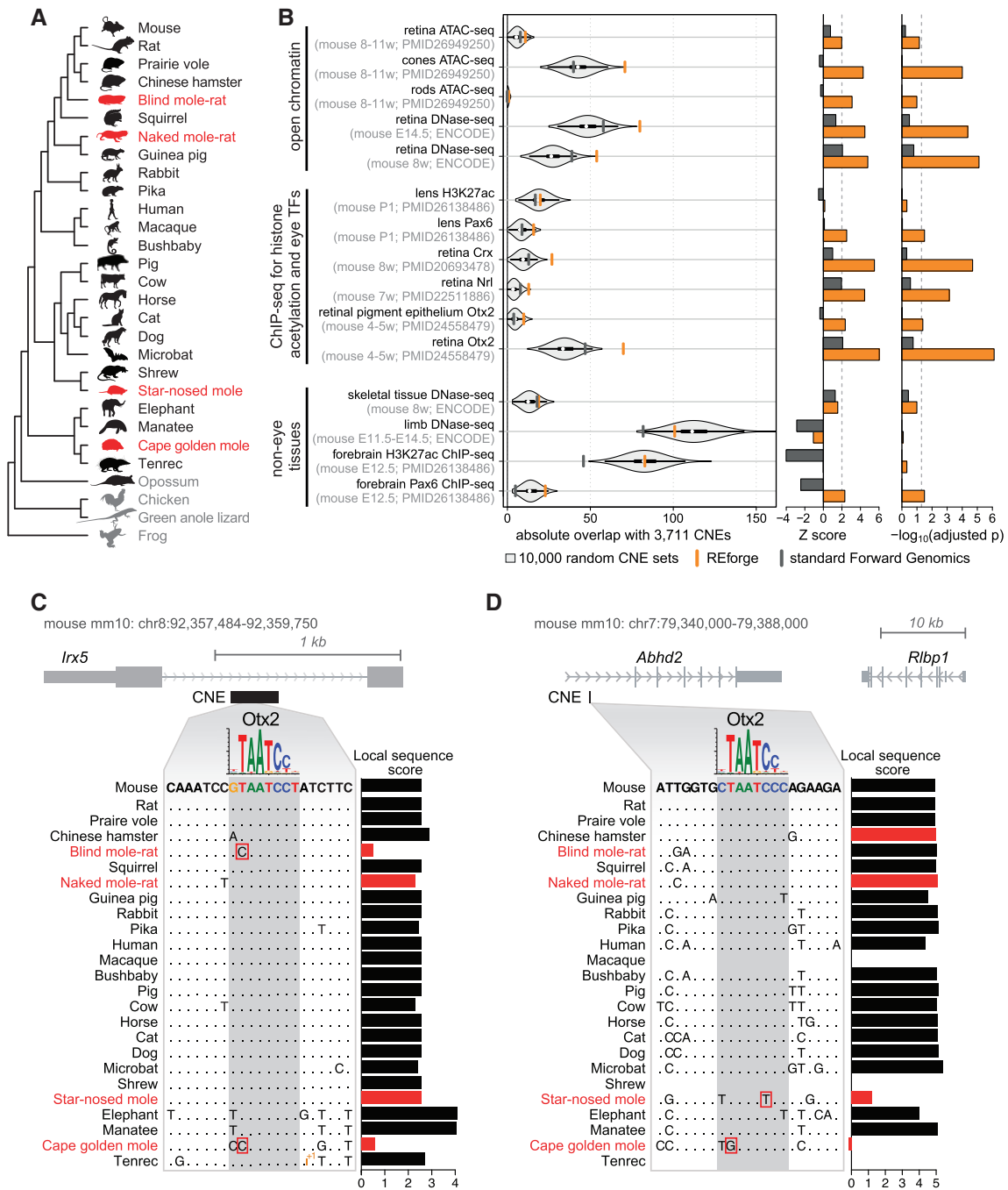
Fig. 3. Genome-wide screen for TFBS divergence in CNEs in subterranean mammals. (*A*) Phylogenetic tree of the species included in the multiple genome alignment. The four independent subterranean lineages that have degenerated eyes are in red, species that are outgroups to eutherian mammals are in gray. (*B*) Diverged CNEs identified with REforge significantly overlap eye-related regulatory data sets. Left: Orange and gray vertical bars compare the observed overlap of the 3,711 top-ranked CNEs identified with REforge and standard Forward Genomics, respectively. The expected overlap was determined by randomly sampling 3,711 CNEs from all CNEs and plotting the overlap of 10,000 such subsets as gray violin plots (thick horizontal line spans the first and the third quartile, white dot is the median). Middle: *Z*-scores measure the number of SDs that the observed overlap is above the random expectation. Right: Benjamini–Hochberg adjusted *P* values obtained with a one-sided Fisher's exact test. (*C* and *D*) Two examples where minor sequence changes destroy a homeobox TF binding site in subterranean mammals. Both CNEs rank highly among the REforge-identified CNEs but are not identified as significantly diverged by standard Forward Genomics, which measures overall sequence divergence. (*C*) A CNE located 1-kb downstream of the *Irx5* transcription start site overlaps a H3K27ac enhancer mark and a region bound by Otx2 in the mouse retina. The convergent T>C change in blind mole-rat and cape golden mole results in the loss of this TFBS, reflected by the low sequence scores. (*D*) A CNE located 44 kb downstream of *Rlbp1* overlaps a cone-specific ATAC-seq peak and a region bound by Otx2 in mouse retina. Independent single base pair substitutions in star-nosed mole and cape golden mole weaken or destroy this TFBS. Since REforge uses Stubb, which does not output the positions of TFBSs, we used MAST ([Bailey and Gribskov 1998](#)) to locate TFBSs in these CNEs.

transcription factors such as Pax6, Crx, Nrl, and Otx2 (Corbo et al. 2010; Hao et al. 2012; Samuel et al. 2014; Sun et al. 2015). We also found a significant overlap with genomic regions marked by accessible chromatin (a hallmark of regulatory activity) in mouse retina tissue and in cone and rod photoreceptor cells (Encode Project Consortium 2012; Mo et al. 2016). This shows that the 3,711 CNEs with binding site divergence in subterranean mammals overlap eye-regulatory data sets significantly more often than expected.

In total, 264 out of the 3,711 CNEs with diverged TFBS overlap eye-specific regulatory regions (supplementary table 3, Supplementary Material online). Subsampling from the publicly available regulatory data sets shows that the overlap with the detected CNEs does not saturate, indicating that additional CNEs may overlap eye-specific regulatory regions not yet sampled (supplementary fig. 8, Supplementary Material online). Yet, these 264 CNEs already highlight a number of regulatory elements that are located near genes with key roles in eye development and function. For example, we detected a CNE located in the first intron of *Irx5* (fig. 3C), which encodes a homeobox TF required for the development of retinal bipolar cells, a type of neuron that transmit signals from cone photoreceptors to ganglion cells (Cheng et al. 2005). Another CNE is located in the second intron of *Rax*, another homeobox TF that is required for retinal progenitor cell proliferation and the formation of the optic cup (Mathers et al. 1997). REforge also detected a CNE near *Eomes*, a T-box TF required for differentiation of retinal ganglion cells and proper development of the optic nerve (Mao et al. 2008). Other CNEs that overlap eye-regulatory regions are located near genes that have important roles in photoreceptors, such as *Crb1*, an adherens junction transmembrane protein required for photoreceptor maintenance (van de Pavert et al. 2004), *Rlbp1*, a gene necessary for efficient rhodopsin regeneration (Saari et al. 2001), or *Neurod1*, a basic helix-loop-helix transcription factor required for photoreceptor survival (Pennesi et al. 2003; Ochocinska et al. 2012). REforge also detected a CNE that overlaps the promoter of *Gnat2*, which encodes a cone-specific subunit of transducin, a G-protein that is essential for visual phototransduction (Chang et al. 2006). Finally, two additional CNEs are located near *Tdrd7*, which encodes an RNA granule component required for normal lens development (Lachke et al. 2011). Notably, mutations in many of these genes are associated with eye diseases in human, ranging from night blindness (*Rlbp1*), total color blindness (*Gnat2*), retinitis pigmentosa (*Neurod1*, *Crb1*), Leber congenital amaurosis (*Crb1*), glaucoma, and cataracts (*Tdrd7*) to severe disorders such as anophthalmia (absence of one or both eyes, *Rax*) (Lotery et al. 2001; Kohl et al. 2002; Voronina et al. 2004; Lequeux et al. 2008; Lachke et al. 2011; Wang et al. 2014).

Next, we compared REforge with standard Forward Genomics on real data. To this end, we applied standard Forward Genomics to all 351,279 CNEs and selected the same number (3,711) of top-ranked CNEs to ensure comparability. The sets of top-ranked 3,711 CNEs identified with REforge and Forward Genomics have an overlap of 153 CNEs (supplementary table 3, Supplementary Material

online), which is 4 times more than the overlap of 39 ($3,711^2/351,279$) CNEs expected by chance. We next assessed which of the two sets of 3,711 CNEs has a higher overlap with eye-regulatory data. In contrast to REforge, the top-ranked 3,711 CNEs identified with standard Forward Genomics are not significantly enriched in these eye-regulatory data sets (fig. 3B). 178 of these 3,711 CNEs overlap eye-related regulatory data sets, which is substantially less than the 264 CNEs that REforge identified. Comparing the top-ranked 9,364 CNEs, which corresponds to a Forward Genomics adjusted P value cutoff of 0.005, corroborates these results (supplementary fig. 9, Supplementary Material online). Overall, REforge also outperforms standard Forward Genomics with respect to enrichments and total overlap with eye-regulatory regions.

REforge's improved sensitivity in comparison to standard Forward Genomics likely reflects its ability to detect small mutations that destroy binding sites for important eye TFs. For example, the CNE near *Irx5* (discussed above) exhibits a T > C mutation in the blind mole-rat and cape golden mole that destroys a putative binding site for Otx2 that is otherwise conserved among mammals (fig. 3C). The CNE associated with *Rlbp1* is another example for the loss of a homeobox TFBS by a single substitution (fig. 3D). Such functionally important sequence changes in an otherwise conserved region are not detectable with standard Forward Genomics.

Since REforge with eye TFs as input should primarily detect CNEs that are active in eye tissues, but not necessarily in other tissues, we investigated the overlap with noneye-related regulatory data sets. To this end, we tested the significance of the overlap of the 3,711 CNEs with published data sets from noneye tissues such as forebrain, skeletal, or limb tissues. We found no significant overlap between the identified CNEs and skeleton or limb data sets. Interestingly, however, we found a significant enrichment with Pax6 bound regions identified by ChIP-seq in forebrain tissue. A potential explanation is that Pax6 is one of the master regulators of eye development and eyes develop from optic vesicles, which originate from forebrain tissue. Furthermore, Pax6 also has a role in optic nerve development, as Pax6 mutations in humans are associated with optic nerve malformations (Azuma et al. 2003). Thus, binding site divergence in some of these overlapping CNEs may be involved in the degenerated optic nerves that characterize the vision-impaired subterranean mammals (Herbin et al. 1995; Catania 1999; Hetling et al. 2005; Nemec et al. 2008). This suggests that the 3,711 CNEs with binding site divergence in subterranean mammals not only overlap regulatory data sets obtained from eye tissue but also from tissue relevant for eye development.

Discussion

Here, we present REforge, a method to associate TFBS divergence with phenotypic differences between species. Like the Forward Genomics branch method (Prudent et al. 2016), REforge makes use of ancestral reconstruction to consider evolutionary changes that happened on each individual branch. In contrast to standard Forward Genomics, REforge

does not measure sequence divergence, but estimates differences in the collective binding affinity of a set of TFs on every branch. This is important since sequence divergence in CREs may not necessarily result in functional divergence, which is conceptually similar to coding regions where synonymous changes contribute to overall sequence divergence but do not alter the encoded protein. Using large simulated data sets, we show that REforge is able to specifically detect those CREs that evolve under no selection to preserve TFBS on the trait-loss branches. While the evolution of real CREs is certainly more complex than what is captured in the simulation, we found that many of the top-ranked CREs consistently correspond to true positives, suggesting that REforge is robust to different scenarios and parameter settings in our simulation. Furthermore, REforge substantially outperforms standard Forward Genomics, both on simulated and on real data, showing that REforge is the superior method when applied to *cis*-regulatory elements.

We applied REforge to screen genome-wide for CRE candidates that exhibit preferential divergence of binding sites for eye-related TFs in subterranean mammals. Out of 351,279 CNEs, REforge detected 3,711 (1.06%) elements. We showed that these CNEs overlap a number of eye-regulatory data sets significantly more often than expected by chance. Our genome-wide screen largely extends a recent study that considered 4,000 CNEs near 20 eye-related TFs and 946 mouse eye enhancers, and successfully identified accelerated sequence (though not necessarily binding site) divergence in subterranean mammals in 29 of these genomic regions (Partha et al. 2017). While the lack of a genome-wide screen precludes a direct comparison to our results, our study uncovered 9 times (264) as many CNEs that overlap eye-specific regulatory elements. Notably, several of these CNEs, described here for the first time, are located near key genes implicated in eye development and human eye disease. These findings suggest that, in addition to a number of lost or diverged eye-related genes (Kim et al. 2011; Emerling and Springer 2014; Fang et al. 2014; Prudent et al. 2016; Partha et al. 2017), several diverged eye-regulatory elements likely contributed to loss of vision and eye degeneration in subterranean mammals. Furthermore, our screen provides evidence for a broader, genome-wide divergence signature of the eye-regulatory landscape in mammals with degenerated eyes.

REforge provides a new tool to discover associations between binding site divergence in regulatory elements and morphological or other phenotypic changes. The discovery of new associations between regulatory and phenotypic differences will benefit from integrating comparative genomics approaches like REforge with high-throughput functional genomics approaches like ATAC-, DNase-, or ChIP-seq. Functional genomics methods typically detect many thousands of regulatory elements in selected tissues, which leaves the challenge of identifying which of these regulatory elements are relevant for a particular phenotypic change. Intersecting such regulatory data with comparative genomics data will help refine the list of relevant genomic regions showing binding site divergence in species where the selected phenotype has changed. Such regulatory elements are promising

candidates for subsequent functional experiments. Facilitated by the rapid increase in the number of sequenced genomes, REforge has broad applicability to detect CRE candidates that could be involved in many other phenotypic differences between sequenced species, and thus help to uncover the genomic basis of nature's phenotypic diversity.

Materials and Methods

REforge Method

Sequence Scores

To compute sequence scores for extant species and reconstructed ancestral sequences, it is necessary that the input genomic regions align between the species. However, it is not required for the genomic regions to be evolutionarily conserved, as REforge can also identify binding site divergence in nonconserved genomic regions when applied to a set of aligning regulatory elements (supplementary fig. 10 and table 4, Supplementary Material online). Given an alignment of a genomic region with reconstructed ancestral sequences and a set of TF motifs, REforge estimates the collective binding affinity of all TFs to the sequence in every extant and ancestral species using Stubb version 2.1 (Sinha et al. 2003, 2006). Stubb makes use of two Hidden Markov models (HMM); one that emits either background sequence or TF binding sites by sampling from one of the given motifs, and one that emits only background sequence. For each HMM, Stubb computes the weighted sum of all paths through the model that emit the given sequence, and outputs the ratio of the scores for the two HMMs. Stubb either optimizes the HMM transition probabilities via expectation maximization or takes them as given. REforge makes use of both options. First, we let Stubb estimate the optimal transition probabilities for the sequence of the common ancestor of all species. Second, these transition probabilities are reused for every sequence that represents a species (extant species or internal node) that descends from the common ancestor. This procedure ensures comparability in the Stubb scores by avoiding fluctuations in the transition probabilities (see below), but can be switched off with parameter *–no_fixed_TP*. To further ensure comparability, we used a fixed predefined set of background sequences as input to Stubb to infer the HMM emission probabilities for the background sequence. Specifically, we generated random sequences and sorted them into 20 bins according to their GC-content. The background sequence belonging to the bin that best matches the input sequence's GC-content is then used as background input to Stubb. The resulting Stubb score reflects the collective affinity of the given TF set.

Although randomly generated sequences should contain TF binding sites only by chance, Stubb scores are consistently positive for such random sequences (supplementary fig. 1, Supplementary Material online). To obtain scores that are on an average zero for random sequences, we created ten randomized sequences by shuffling the nucleotides in the input sequence and subtracted the average Stubb score of these ten sequences from the score of the input sequence. The resulting score is called the "sequence score" (supplementary fig. 2, Supplementary Material online).

Branch Scores

Since species are related by the phylogeny, the sequence scores of extant or ancestral species (terminal or internal nodes in the tree) are not independent and thus cannot be directly compared. Therefore, we adapt the branch method from the Forward Genomics framework (Prudent et al. 2016) that compares evolutionary changes between the branches in the phylogenetic tree instead of comparing changes in the nodes of the tree. We compute a "branch score" as the difference of the sequence scores of the end and start node of this branch. Positive branch scores represent gain or strengthening of binding sites and negative scores indicate the weakening or loss of binding sites along the branch. Since every branch represents an independent evolutionary trajectory, the branch scores are phylogenetically independent and can be directly compared. Instead of directly determining the number of sequence changes that happened on this branch, branch scores measure relative changes in TFBSs and thus are not dependent on the length of the branch.

Two different scenarios can explain a branch score of ∼0. First, if both start and end node of the branch have positive sequence score, then TF binding sites are present and are largely preserved (though not necessarily at the same position). Second, if both nodes have sequence scores ≤0, then no significant TF binding motifs are present in either sequence and we did not consider such uninformative branch scores.

Association Test

To test if there is an association between the branch scores and a repeatedly evolved binary trait, we use Dollo parsimony to infer the phenotype of all ancestors. Alternatively, Maximum Likelihood approaches can be used for that purpose. Then, we classify each branch as either "trait-preserving" or "trait-loss," depending on the trait state of the branch end node. If trait loss is associated with the progressive divergence or loss of TF motifs along the trait-loss branches, we expect that these branches preferentially have negative branch scores, in contrast to trait-preserving branches, where TF binding affinity is under selection resulting in positive branch scores or scores ∼0. To test if trait-loss branches have significantly lower branch scores than trait-preserving branches, we used the significance of a positive Pearson correlation coefficient. It should be noted that while we here only consider an association between trait-loss branches and negative branch scores, the association principle is general and can also be used to detect associations between a set of branches and positive branch scores that reflect gains of TF motifs. The final output of REforge is the P value of a given genomic region.

Simulating the Evolution of CREs

We simulated the evolution of CREs to obtain data sets where the ground truth (which CREs are and are not associated with the phenotype) is known. In this simulation, we used a set of five "phenotype-relevant TFs" that have motifs which are sufficiently different from each other (supplementary fig. 4A, Supplementary Material online). We used a phylogeny

of 20 species (fig. 2A) and selected three independent species as trait-loss species.

As ancestral CREs, we used 200-bp sequences that were randomly generated according to a uniform base distribution and placed five consensus binding sites for randomly selected TFs at random but nonoverlapping positions. For each constructed sequence, we used GEMSTAT (He et al. 2010), a method that predicts regulatory activity (expression pattern) from the TFBSs in the sequence, to measure the fitness of the sequence. A fitness of 1 corresponds to a perfect match between the regulatory activity predicted for the sequence and the target regulatory activity. The target regulatory activity of a CRE was defined as resulting in 100% expression level in a single tissue. We simulated a simple regulatory logic where the five TFs have equal concentrations levels in this tissue and are independent activators with equal strength, thus each TF contributes equally to regulatory activity. To ensure that the ancestral sequence at the start of the simulation has the appropriate regulatory activity, we discarded all sequences with a fitness of <0.85. In total, we generated 10,200 different ancestral sequences. Starting with an ancestral sequence, the evolution of each phylogenetic branch was simulated by a successive application of PEBCRES (Duque et al. 2014), a discrete-time Wright–Fisher simulation with a fixed size population. The mutation parameters of PEBCRES were set to *mutation_rate 1e-04*, *substitution_probability 0.95*, *insertion_probability 0.5*, and *tandem_repeat_probability 0.2*. For every branch, we ran PEBCRES with a number of iterations (parameter "*num_generations*") such that the total mutation rate equals the branch length (number of substitutions per site). For example, a branch length of 0.05 substitutions per site corresponds to 500 iterations. Starting from a single ancestral sequence, PEBCRES simulates the evolution of a population of 50 sequences to obtain the final population representing an internal node in the tree. To independently evolve this population along the two branches that descend from this internal node, we modified the PEBCRES source code to start these two independent evolution runs with the set of 50 sequences obtained for this node. This process was repeated until we obtained the 50-sequence population of every internal node and every extant species. Then, for each node, we selected the sequence with the median fitness to obtain a single sequence that represents each node in the tree. As a result, all ancestral states are known after simulating the evolution of a CRE.

For 10,000 of the ancestral sequences, we evolved every branch under selection to preserve the target regulatory activity (high fitness) using the selection parameters $D\_max = 1$, $selectionExp = 2$, $selectionScale = 100$, and $selectionCoeff = 0.1$. These CREs are not associated with trait loss and thus correspond to negatives (background CRE set 1). To simulate 200 foreground CREs that are associated with the trait loss and thus evolve neutrally after the trait was lost, we split the trait-loss branch into two parts. Along the first part, the CRE evolved under selection. Along the second part, which consisted of a branch length of 0.09, 0.06, or 0.03 neutral substitutions per site, the CRE evolved neutrally by setting $selectionCoeff = 0$, which removes the influence of fitness

during the selection step in the Wright–Fisher model. For comparison, 0.09 subs. per site corresponds to the divergence of rat from the rat–mouse ancestor. The final data set consists of 10,200 CREs, 200 (1.96%) of which are associated to trait-loss (positives). We always used the known (simulated) ancestral sequences, except for the tests with the 11-species phylogeny, for which we reconstructed ancestral sequences with PRANK (Loytynoja and Goldman 2008) (parameters "*-once -gaprate = 0.05 -gapext = 0.2 -termgap –showanc*").

To explore robustness of REforge to different backgrounds, we generated an additional set of 5,000 CREs (background CRE set 2) that are active in an unrelated tissue and whose regulatory activity is controlled by five different activator TFs. To this end, we chose five TFs having motifs that are sufficiently different from the five previously chosen motifs using a TomTom (Gupta et al. 2007) similarity score cutoff of 0.01 (supplementary fig. 4B, Supplementary Material online) and repeated the above-described steps to generate a negative CRE set. We also tested the robustness of REforge to variation in the sets of input TF motifs by selecting three sets of 15 additional TFs with motifs that are different from the motifs of the five activator TFs (TomTom similarity score cutoff of 0.01, supplementary table 1, Supplementary Material online).

In order to create three sets of 200 pleiotropic foreground CREs (one set for each trait-loss age), we simulated CRE evolution under a target regulatory activity of 100% expression level in two different tissues. While expression in the first tissue is controlled by the same five TFs as before, expression in the second tissue is controlled by a second set of five activator TFs with equal strength and equal concentrations levels (supplementary fig. 4B, Supplementary Material online). We assumed that the first tissue is related to the lost trait, while the second tissue is not. Therefore, after trait loss, we redefined the target regulatory activity of the CRE to 100% expression level in only the second tissue. This differs from the tissue-specific foreground CREs above, where we removed any selection by setting *selectionCoeff = 0*.

Testing REforge Parameters

We used the simulated data to test and optimize various parameters of REforge. First, we compared the performance of different methods for quantifying the significance of the association between trait and branch scores. As shown in supplementary figure 3, Supplementary Material online, Pearson correlation outperforms several other methods such as ranking CREs by the significance of a positive Spearman correlation coefficient, the significance of a Wilcoxon test or *t*-test, or the upper bound of the 95% confidence interval obtained with a Wilcoxon or *t*-test.

Second, we tested the effect of estimating the optimal Stubb-HMM transition probabilities for the sequence of the common ancestor and then reusing the same transition probabilities for every descendant species. As shown in supplementary figure 11, Supplementary Material online, reusing the optimal ancestral transition probabilities results in a better performance, since it avoids fluctuations in the transition probabilities that would arise if they were optimized for each sequence separately. Reusing the optimal ancestral transition

probabilities has another beneficial side effect. In case, our assumption that the ancestral CRE is important for the given trait and thus exhibits TFBS is violated, the optimal transition probabilities for the sequence that represents the common ancestor are zero. Consequently, all sequences scores and thus all branch scores will be zero, resulting in a *P* value of 1. Therefore, by default, REforge avoids computing sequence scores for such CREs to save runtime.

Detecting CREs Associated with Eye Degeneration in Mammals

The multiple genome alignment was created as previously described in (Prudent et al. 2016). Briefly, we used the mouse mm10 genome assembly as the reference and applied the lastz/chain/net pipeline (Kent et al. 2003; Harris 2007) to obtain pairwise alignments to 25 mammals and 3 outgroups (lizard, chicken, and frog) and Multiz (Blanchette et al. 2004) to build a multiple alignment. To obtain CNEs, we first identified evolutionarily conserved elements with PhastCons (parameters "*expected-length = 45, target-coverage = 0.3 rho = 0.3*") (Siepel et al. 2005) and GERP (default parameters) (Davydov et al. 2010) in this multiple alignment and then excluded all conserved parts that overlap annotated coding exons in mouse. We only considered CNEs that are ≥30 bp long and provide sequence information for all four subterranean species and at least 15 of all species in total. This resulted in 351,279 CNEs. For each CNE, we used the phylogeny-aware PRANK method (Loytynoja and Goldman 2008) (parameters "*-keep -showtree -showanc -prunetree -seed = 10*") to align the sequences of all species and to reconstruct all ancestral sequences (internal nodes in the tree).

We obtained eye-related transcription factors from a recent review (Cvekl and Mitton 2010) and used the TRANSFAC (Matys et al. 2006), JASPAR (Mathelier et al. 2014), and UniPROBE (Hume et al. 2015) motif databases to assign motifs to these TFs. In case, several distinct (primary/secondary) motifs describe the binding preference of a TF, we included all motifs. This resulted in a total of 28 motifs for 19 different TFs (supplementary table 2, Supplementary Material online). Using the 28 motifs as input, we applied REforge to all 351,279 CNEs using a scoring window size of 200, ancestrally fixed transition probabilities and ancestral CNE filter on the ancestor of eutherian mammals. To correct for multiple testing, we applied the Benjamini–Hochberg procedure and used an adjusted *P* value cut-off of 0.1, which resulted in 3,711 CNEs that are associated with eye degeneration. Repeating the same analysis with only one representative motif for each of the 19 eye-related TFs (marked in supplementary table 2, Supplementary Material online) resulted in 3,392 CNEs that are also enriched for overlap with eye-regulatory data sets (supplementary fig. 7, Supplementary Material online).

To compare REforge to the standard Forward Genomics on real data, we applied the branch method with default parameters (Prudent et al. 2016) to all 351,279 CNEs. We selected the 3,711 top-ranked CNEs to compare both methods on a data set of the same size. To also compare both methods to random CNE sets, we randomly selected 10,000

sets of 3,711 CNEs each out of the total 351,279 CNEs. To exclude any bias in comparing REforge to standard Forward Genomics, we also selected the 9,364 top-ranked CNEs that correspond to a Forward Genomics adjusted P value of <0.005 and compared this set to the same number of top-ranked REforge CNEs (supplementary fig. 9, Supplementary Material online).

To test if the CNEs identified with REforge or standard Forward Genomics significantly overlap eye-specific regulatory elements, we compiled a comprehensive list of publicly available eye-regulatory data sets obtained from ChIP-seq, DNase-seq, and ATAC-seq experiments, as previously described (Roscito et al. 2018). These include ChIP-seq binding sites for the transcription factors Crx (Corbo et al. 2010) and Nrl (Hao et al. 2012) in retina of adult mice, for Otx2 (Samuel et al. 2014) in retina of 1-month-old mice, and for Pax6 (Sun et al. 2015) in lenses of newborn (P1) mice. We also included open chromatin regions identified with DNase-seq in both adult and embryonic E14.5 retinas (Encode Project Consortium 2012), and with ATAC-seq in adult retina, cone, and rod photoreceptors (Mo et al. 2016).

To obtain discrete regulatory regions, we processed each data set as follows. Otx2 ChIP-seq peaks were kindly provided by Alexander Samuel (Samuel et al. 2014) and Nrl peaks were kindly provided by Anand Swaroop (Hao et al. 2012). For Otx2, we defined retina- and retinal pigment epithelium (RPE)-specific peaks by selecting those peaks which are unique to each tissue (*bedtools intersect −v*; 14,015 retina-specific peaks and 1,509 RPE-specific peaks). For the Crx data set, we selected the peaks which were supported by at least 40 sequencing reads (2,974 peaks). To obtain lens-specific Pax6 and H3K27ac peaks, we used the forebrain samples from the same study to select peaks marked by Pax6 and H3K27ac in lens but not in the respective forebrain samples (*bedtools intersect −v*; 2,570 lens-specific Pax6 peaks and 6,390 lens-specific H3K27ac peaks). Similarly, we obtained forebrain-specific Pax6 and H3K27ac peaks by selecting those peaks unique to forebrain. To obtain adult retina-specific DNaseI peaks, we downloaded, in addition to the retina peaks, several ENCODE peak sets from 8-week-old mice (whole brain, cerebellum, heart, kidney, liver, lung, skeletal tissue, telencephalon). For each tissue, we combined replicates to obtain peaks consistently found in at least two of the replicates. Next, we combined all nonretina peaks and subtracted them from the retina DNaseI peaks (*bedtools intersect −v*; 12,763 retina-specific DNaseI peaks). We applied the same procedure for ENCODE E14.5 samples, that is, subtracted nonretina peak sets obtained from E11.5 and E14.5 mice embryos (facial prominence, whole brain, forebrain, midbrain, hindbrain, neural tube, whole limb, forelimb, and hindlimb) from retina peaks to obtain 23,933 retina-specific peaks. We also obtained peaks specific to each nonretina tissues, adult and embryonic, following the same procedure. Finally, for the ATAC-seq peak sets, we ran DiffBind (*bFullLibrary Size=FALSE, bTagwise=FALSE, method=EDGER*) (Ross-Innes et al. 2012) with the adult wild-type retina, cones and rods samples to obtain 254 rod-, 20,335 cone-, and 1,507 retina-specific peaks.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

All simulated CREs (10,000/5,000 background CREs, 200 tissue-specific, and 200 pleiotropic foreground CREs for the three trait-loss time points) and the TF motifs are available at https://bds.mpi-cbg.de/hillerlab/REforge/, last accessed October 11, 2018.

Code Availability

The REforge source code is available at https://github.com/hillerlab/REforge, last accessed October 11, 2018.

Acknowledgments

References

Azuma N, Yamaguchi Y, Handa H, Tadokoro K, Asaka A, Kawase E, Yamada M. 2003. Mutations of the PAX6 gene detected in patients with a variety of optic-nerve malformations. *Am J Hum Genet.* 72(6):1565–1570.

Bailey TL, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14(1):48–54.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* 14(4):708–715.

Booker BM, Friedrich T, Mason MK, VanderMeer JE, Zhao J, Eckalbar WL, Logan M, Illing N, Pollard KS, Ahituv N. 2016. Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus. *PLoS Genet.* 12(3):e1005738.

Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. 2013. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci.* 368(1632):20130025.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3(7):e245.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

Catania KC. 1999. A nose that looks like a hand and acts like an eye: the unusual mechanosensory system of the star-nosed mole. *J Comp Physiol A* 185(4):367–372.

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327(5963):302–305.

Chang B, Dacey MS, Hawes NL, Hitchcock PF, Milam AH, Atmaca-Sonmez P, Nusinowitz S, Heckenlively JR. 2006. Cone photoreceptor function loss-3, a novel mouse model of achromatopsia due to a mutation in Gnat2. *Invest Ophthalmol Vis Sci.* 47(11):5017–5021.

Cheng CW, Chow RL, Lebel M, Sakuma R, Cheung HO, Thanabalasingham V, Zhang X, Bruneau BG, Birch DG, Hui CC, et al. 2005. The Iroquois homeobox gene, Irx5, is required for retinal cone bipolar cell development. *Dev Biol.* 287(1):48–60.

Corbo JC, Lawrence KA, Karlstetter M, Myers CA, Abdelaziz M, Dirkes W, Weigelt K, Seifert M, Benes V, Fritsche LG, et al. 2010. CRX ChIP-seq

reveals the cis-regulatory architecture of mouse photoreceptors. *Genome Res.* 20(11):1512–1525.

Cretekos CJ, Wang Y, Green ED, Martin JF, Rasweiler JJt, Behringer RR. 2008. Regulatory divergence modifies limb length between mammals. *Genes Dev.* 22(2):141–151.

Cvekl A, Mitton KP. 2010. Epigenetic regulatory mechanisms in vertebrate eye development and disease. *Heredity (Edinb)* 105(1):135–151.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 6(12):e1001025.

Duque T, Samee MA, Kazemian M, Pham HN, Brodsky MH, Sinha S. 2014. Simulations of enhancer evolution provide mechanistic insights into gene regulation. *Mol Biol Evol.* 31(1):184–200.

Emerling CA, Springer MS. 2014. Eyes underground: regression of visual protein networks in subterranean mammals. *Mol Phylogenet Evol.* 78:260–270.

Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.

Fang X, Nevo E, Han L, Levanon EY, Zhao J, Avivi A, Larkin D, Jiang X, Feranchuk S, Zhu Y, et al. 2014. Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. *Nat Commun.* 5:3966.

Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. 2015. Suboptimization of developmental enhancers. *Science* 350(6258):325–328.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol.* 8(2):R24.

Hao H, Kim DS, Klocke B, Johnson KR, Cui K, Gotoh N, Zang C, Gregorski J, Gieser L, Peng W, et al. 2012. Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. *PLoS Genet.* 8(4):e1002649.

Harris RS. 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

He X, Samee MA, Blatti C, Sinha S. 2010. Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput Biol.* 6(9):e1000935.

Herbin M, Rio JP, Reperant J, Cooper HM, Nevo E, Lemire M. 1995. Ultrastructural study of the optic nerve in blind mole-rats (Spalacidae, Spalax). *Vis Neurosci.* 12(2):253–261.

Hetling JR, Baig-Silva MS, Comer CM, Pardue MT, Samaan DY, Qtaishat NM, Pepperberg DR, Park TJ. 2005. Features of visual function in the naked mole-rat *Heterocephalus glaber*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol.* 191(4):317–330.

Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. 2012. A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep.* 2(4):817–823.

Huang W, Nevins JR, Ohler U. 2007. Phylogenetic simulation of promoter evolution: estimation and modeling of binding site turnover events and assessment of their impact on alignment tools. *Genome Biol.* 8(10):R225.

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. 2015. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 43(Database issue):D117–D122.

Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, Myers RM, Kingsley DM. 2016. Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell* 164(1–2):45–56.

Infante CR, Mihala AG, Park S, Wang JS, Johnson KK, Lauderdale JD, Menke DB. 2015. Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes. *Dev Cell* 35(1):107–119.

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell* 132(5):783–793.

Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A, Fischman BJ, et al. 2015. Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348(6239):1139–1143.

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A.* 100(20):11484–11489.

Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479(7372):223–227.

Kohl S, Baumann B, Rosenberg T, Kellner U, Lorenz B, Vadala M, Jacobson SG, Wissinger B. 2002. Mutations in the cone photoreceptor G-protein alpha-subunit gene GNAT2 in patients with achromatopsia. *Am J Hum Genet.* 71(2):422–425.

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissieres V, Pickle CS, Plajzer-Frick I, Lee EA, et al. 2016. Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167(3):633–642 e611.

Lachke SA, Alkuraya FS, Kneeland SC, Ohn T, Aboukhalil A, Howell GR, Saadi I, Cavallesco R, Yue Y, Tsai AC, et al. 2011. Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. *Science* 331(6024):1571–1576.

Leal F, Cohn MJ. 2016. Loss and re-emergence of legs in snakes by modular evolution of sonic hedgehog and HOXD enhancers. *Curr Biol.* 26(21):2966–2973.

Lequeux L, Rio M, Vigouroux A, Titeux M, Etchevers H, Malecaze F, Chassaing N, Calvas P. 2008. Confirmation of RAX gene involvement in human anophthalmia. *Clin Genet.* 74(4):392–395.

Lotery AJ, Jacobson SG, Fishman GA, Weleber RG, Fulton AB, Namperumalsamy P, Heon E, Levin AV, Grover S, Rosenow JR, et al. 2001. Mutations in the CRB1 gene cause Leber congenital amaurosis. *Arch Ophthalmol.* 119(3):415–420.

Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320(5883):1632–1635.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45(D1):D896–D901.

Mao CA, Kiyama T, Pan P, Furuta Y, Hadjantonakis AK, Klein WH. 2008. Eomesodermin, a target gene of Pou4f2, is required for retinal ganglion cell and optic nerve development in the mouse. *Development* 135(2):271–280.

Marcovitz A, Jia R, Bejerano G. 2016. "Reverse Genomics" predicts function of human conserved noncoding elements. *Mol Biol Evol.* 33(5):1358–1369.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, et al. 2014. 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42(Database issue):D142–D147.

Mathers PH, Grinberg A, Mahon KA, Jamrich M. 1997. The Rx homeobox gene is essential for vertebrate eye development. *Nature* 387(6633):603–607.

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, et al. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34(Database issue):D108–D110.

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471(7337):216–219.

Mo A, Luo C, Davis FP, Mukamel EA, Henry GL, Nery JR, Urich MA, Picard S, Lister R, Eddy SR, et al. 2016. Epigenomic landscapes of retinal rods and cones. *Elife* 5:e11613.

Nemec P, Cvekova P, Benada O, Wielkopolska E, Olkowicz S, Turlejski K, Burda H, Bennett NC, Peichl L. 2008. The visual system in subterranean African mole-rats (Rodentia, Bathyergidae): retina, subcortical visual nuclei and primary visual cortex. *Brain Res Bull.* 75(2–4):356–364.

Nevo E. 1979. Adaptive convergence and divergence of subterranean mammals. *Annu Rev Ecol Evol Syst.* 10:247–267.

O'Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. 2015. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *Elife* 4:e05290.

Ochocinska MJ, Munoz EM, Veleri S, Weller JL, Coon SL, Pozdeyev N, Iuvone PM, Goebbels S, Furukawa T, Klein DC. 2012. NeuroD1 is required for survival of photoreceptors but not pinealocytes: results from targeted gene deletion studies. *J Neurochem.* 123(1):44–59.

Otto W, Stadler PF, Lopez-Giraldez F, Townsend JP, Lynch VJ, Wagner GP. 2009. Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. *Genome Biol Evol.* 1:85–98.

Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife* 6:e25884.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444(7118):499–502.

Pennesi ME, Cho JH, Yang Z, Wu SH, Zhang J, Wu SM, Tsai MJ. 2003. BETA2/NeuroD1 null mice: a new model for transcription factor-dependent photoreceptor degeneration. *J Neurosci.* 23(2):453–461.

Prudent X, Parra G, Schwede P, Roscito JG, Hiller M. 2016. Controlling for phylogenetic relatedness and evolutionary rates improves the discovery of associations between species' phenotypic and genomic differences. *Mol Biol Evol.* 33(8):2135–2150.

Roscito JG, Sameith K, Parra G, Langer BE, Petzold A, Moebius C, Bickle M, Rodrigues MT, Hiller H. 2018. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. Nature Communications, in press.

Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481(7381):389–393.

Saari JC, Nawrot M, Kennedy BN, Garwin GG, Hurley JB, Huang J, Possin DE, Crabb JW. 2001. Visual cycle impairment in cellular retinaldehyde binding protein (CRALBP) knockout mice results in delayed dark adaptation. *Neuron* 29(3):739–748.

Saito T, Rehmsmeier M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10(3):e0118432.

Samuel A, Housset M, Fant B, Lamonerie T. 2014. Otx2 ChIP-seq reveals unique and redundant functions in the mature mouse retina. *PLoS One* 9(2):e89110.

Sanyal S, Jansen HG, de Grip WJ, Nevo E, de Jong WW. 1990. The eye of the blind mole rat, *Spalax ehrenbergi*. Rudiment with hidden function? *Invest Ophthalmol Vis Sci.* 31(7):1398–1404.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–1050.

Sinha S, Liang Y, Siggia E. 2006. Stubb: a program for discovery and analysis of cis-regulatory modules. *Nucleic Acids Res.* 34(Web Server issue):W555–W559.

Sinha S, van Nimwegen E, Siggia ED. 2003. Stubb: a probabilistic method to detect regulatory modules. *Bioinformatics* 19(Suppl 1):i292–i301.

Sun J, Rockowitz S, Xie Q, Ashery-Padan R, Zheng D, Cvekl A. 2015. Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. *Nucleic Acids Res.* 43(14):6827–6846.

van de Pavert SA, Kantardzhieva A, Malysheva A, Meuleman J, Versteeg I, Levelt C, Klooster J, Geiger S, Seeliger MW, Rashbass P, et al. 2004. Crumbs homologue 1 is required for maintenance of photoreceptor cell polarization and adhesion during light exposure. *J Cell Sci.* 117(Pt 18):4169–4177.

Villar D, Flicek P, Odom DT. 2014. Evolution of transcription factor binding in metazoans—mechanisms and functional implications. *Nat Rev Genet.* 15(4):221–233.

Voronina VA, Kozhemyakina EA, O'Kernick CM, Kahn ND, Wenger SL, Linberg JV, Schneider AS, Mathers PH. 2004. Mutations in the human RAX homeobox gene in a patient with anophthalmia and sclerocornea. *Hum Mol Genet.* 13(3):315–322.

Wang F, Li H, Xu M, Li H, Zhao L, Yang L, Zaneveld JE, Wang K, Li Y, Sui R, et al. 2014. A homozygous missense mutation in NEUROD1 is associated with nonsyndromic autosomal recessive retinitis pigmentosa. *Invest Ophthalmol Vis Sci.* 56(1):150–155.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3(1):e7.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8(3):206–216.