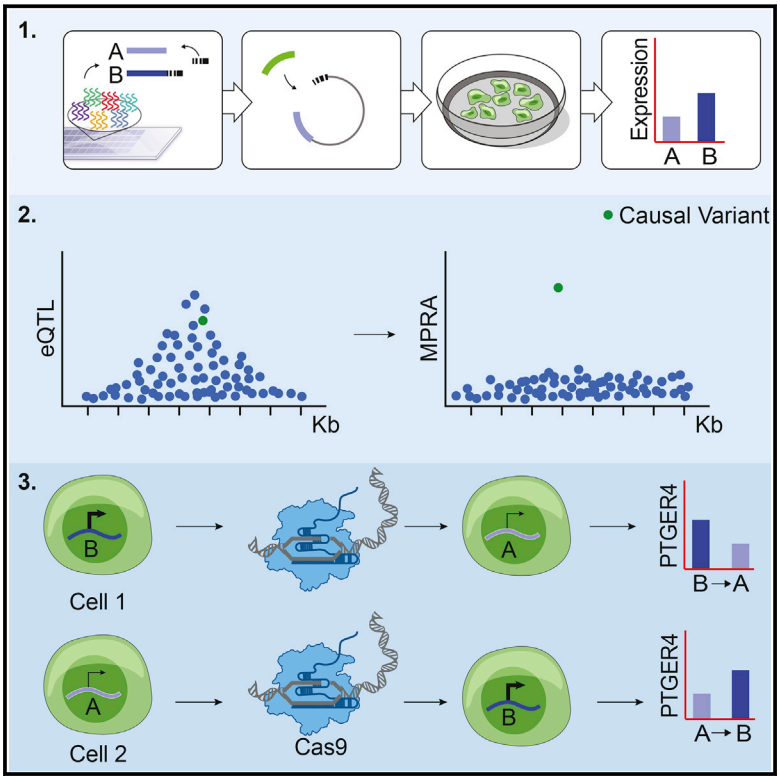
Resource

Cell

Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay

Graphical Abstract



Authors

Ryan Tewhey, Dylan Kotliar, Daniel S. Park, ..., Eric S. Lander, Stephen F. Schaffner, Pardis C. Sabeti

Correspondence

rtewhey@broadinstitute.org (R.T.), pardis@broadinstitute.org (P.C.S.)

In Brief

A massively parallel reporter assay analyzes thousands of human polymorphisms to identify alleles that impact gene expression, providing a tool with which to move from disease-associated GWAS hits to the identification of functional variants.

Highlights

- A new version of MPRA with greater throughput and sensitivity
- Evaluation of 32,373 variants associated with eQTLs in lymphoblastoid cell lines
- 842 variants showed differential gene expression between alleles
- Use of CRISPR/cas9 to identify a distal eQTL causal allele for *PTGER4*

Accession Numbers

GSE75661

Tewhey et al., 2016, Cell *165*, 1519–1529
June 2, 2016
http://dx.doi.org/10.1016/j.cell.2016.04.027

CellPress

Resource

Cell

Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay

Ryan Tewhey,[1,2,*] Dylan Kotliar,[1,2] Daniel S. Park,[2] Brandon Liu,[2] Sarah Winnicki,[2] Steven K. Reilly,[1,2] Kristian G. Andersen,[1,2] Tarjei S. Mikkelsen,[2] Eric S. Lander,[2] Stephen F. Schaffner,[2] and Pardis C. Sabeti[1,2,*]

[1]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

[2]Broad Institute, Cambridge, MA 02142, USA

*Correspondence: rtewhey@broadinstitute.org (R.T.), pardis@broadinstitute.org (P.C.S.)

http://dx.doi.org/10.1016/j.cell.2016.04.027

SUMMARY

Although studies have identified hundreds of loci associated with human traits and diseases, pinpointing causal alleles remains difficult, particularly for non-coding variants. To address this challenge, we adapted the massively parallel reporter assay (MPRA) to identify variants that directly modulate gene expression. We applied it to 32,373 variants from 3,642 cis-expression quantitative trait loci and control regions. Detection by MPRA was strongly correlated with measures of regulatory function. We demonstrate MPRA's capabilities for pinpointing causal alleles, using it to identify 842 variants showing differential expression between alleles, including 53 well-annotated variants associated with diseases and traits. We investigated one in detail, a risk allele for ankylosing spondylitis, and provide direct evidence of a non-coding variant that alters expression of the prostaglandin EP_4 receptor. These results create a resource of concrete leads and illustrate the promise of this approach for comprehensively interrogating how non-coding polymorphism shapes human biology.

INTRODUCTION

The genomic era has enormously increased our knowledge of human genetic variation, but our understanding of the functional consequences of that variation has not kept pace (Cooper and Shendure, 2011). Although genome-wide association studies (GWAS) and whole-genome scans for natural selection have identified numerous loci linked to human traits and diseases, correlation between nearby polymorphisms (linkage disequilibrium or LD) within individual associations often leaves dozens to hundreds of potential causal variants to be interrogated (Grossman et al., 2013; Schaub et al., 2012). Mounting evidence suggests that at the majority of these loci, the causal variant(s) is a non-coding regulatory change rather than an amino acid substitution (Farh et al., 2015; Maurano et al., 2012). Indeed, regulatory changes drive some of the best understood examples of phenotypic diversity and adaptive evolution (Claussnitzer et al., 2015; Musunuru et al., 2010; Tishkoff et al., 2007). Therefore, it is critical that we be able to test whether a variant affects gene regulation.

Current approaches for measuring a variant's effect on gene expression fall into two categories, each with its own limitation. Indirect methods, such as whole-genome epigenetic assays, can only identify the broader regulatory state of a region, not necessarily the effect of a particular variant (Andersson et al., 2014; ENCODE Project Consortium, 2012; Kasowski et al., 2013; McVicker et al., 2013). Direct methods, ones that measure the impact of individual alleles in an episomal or native context on gene expression, are currently low throughput and require substantial resources for comprehensive evaluation of a region.

We adopted the massively parallel reporter assay (MPRA) as a solution and modified it so that we could carry out large-scale, sensitive, and direct testing of potential regulatory variants. This assay is based upon the well-established reporter gene assay, in which a vector containing a reporter gene (e.g., luciferase or green fluorescent protein [GFP]), a promoter, and a potential regulatory sequence is inserted into a plasmid, which is transfected into a cell; sequences that regulate gene expression then alter the amount of luciferase/GFP expressed (Arnold et al., 2013; Melnikov et al., 2012; Ow et al., 1986; Patwardhan et al., 2012; Vockley et al., 2015; Kwasnieski et al., 2014). Through the use of unique barcodes in the 3′ UTR of the reporter to differentiate expression of individual oligos, MPRA can test many different sequences simultaneously, and it has been shown to reproducibly detect segments of the genome that change expression levels (Kheradpour et al., 2013; Mogno et al., 2013). We aimed to incorporate single-nucleotide and small-insertion/deletion polymorphisms (referred to below as single-nucleotide variants or SNVs) into these assays to see whether we could detect subtle differences in how each allele drives expression. Because we used only a minimal promoter, with very low baseline expression, in this iteration of the assay, we intended it primarily as a test of regulatory sequence that increases (i.e., enhancers and promoters), rather than decreases, expression; the latter will be difficult to detect because baseline expression is already low.

Ideally, one would test the assay for sensitivity and specificity by applying it to a set of "gold-standard" variants previously identified as expression quantitative trait loci (eQTLs) that act on enhancer and promoter elements. However, there is a dearth of such known variants. As the best available alternative, we studied