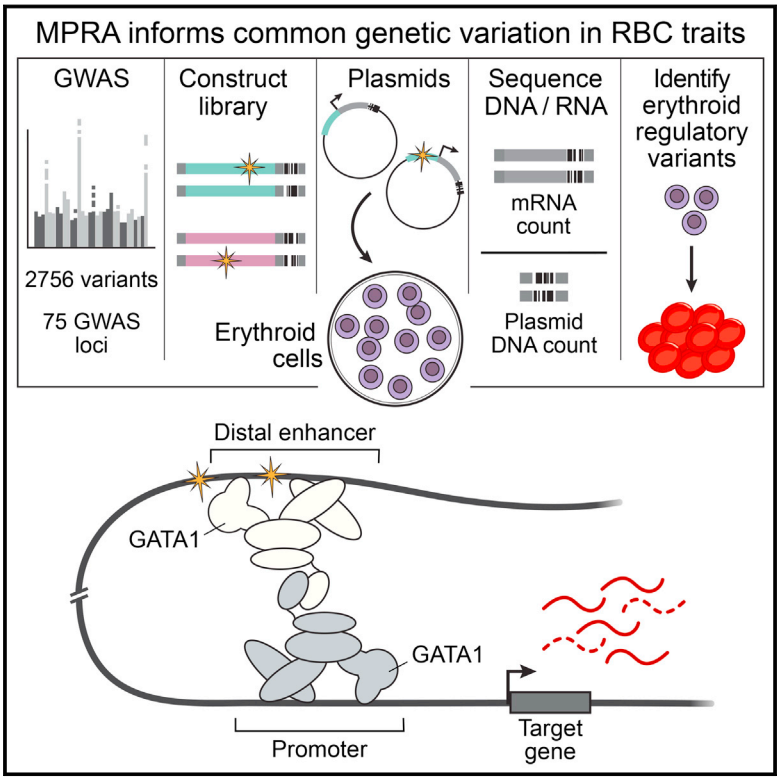
Resource

Cell

Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits

Graphical Abstract



Authors

Jacob C. Ulirsch, Satish K. Nandakumar, Li Wang, ..., Ron Do, Tarjei S. Mikkelsen, Vijay G. Sankaran

Correspondence

sankaran@broadinstitute.org

In Brief

A cost-effective, scalable, and allele-specific assay is used to systematically screen for functional non-coding genetic variation affecting red blood cell traits.

Highlights

- A massively parallel reporter assay was developed to screen for functional variation
- Variants identified by this assay are enriched for orthogonal measures of function
- Functional GWAS variants alter activity of master transcription factors
- The target gene *RBM38* was linked to its GWAS phenotype and regulates mRNA splicing

Ulirsch et al., 2016, Cell *165*, 1530–1545
June 2, 2016
http://dx.doi.org/10.1016/j.cell.2016.04.048

CellPress

Cell

Resource

Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits

Jacob C. Ulirsch,[1,2,6] Satish K. Nandakumar,[1,2,6] Li Wang,[2] Felix C. Giani,[1,2,3] Xiaolan Zhang,[2] Peter Rogov,[2] Alexandre Melnikov,[2] Patrick McDonel,[2] Ron Do,[4] Tarjei S. Mikkelsen,[2,5] and Vijay G. Sankaran[1,2,5,*]

[1]Division of Hematology/Oncology, The Manton Center for Orphan Disease Research, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[3]Charité-Universitätsmedizin Berlin, Berlin 10117, Germany

[4]Department of Genetics and Genomic Sciences and The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[5]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

[6]Co-first author

*Correspondence: sankaran@broadinstitute.org

http://dx.doi.org/10.1016/j.cell.2016.04.048

SUMMARY

Genome-wide association studies (GWAS) have successfully identified thousands of associations between common genetic variants and human disease phenotypes, but the majority of these variants are non-coding, often requiring genetic fine-mapping, epigenomic profiling, and individual reporter assays to delineate potential causal variants. We employ a massively parallel reporter assay (MPRA) to simultaneously screen 2,756 variants in strong linkage disequilibrium with 75 sentinel variants associated with red blood cell traits. We show that this assay identifies elements with endogenous erythroid regulatory activity. Across 23 sentinel variants, we conservatively identified 32 MPRA functional variants (MFVs). We used targeted genome editing to demonstrate endogenous enhancer activity across 3 MFVs that predominantly affect the transcription of *SMIM1*, *RBM38*, and *CD164*. Functional follow-up of *RBM38* delineates a key role for this gene in the alternative splicing program occurring during terminal erythropoiesis. Finally, we provide evidence for how common GWAS-nominated variants can disrupt cell-type-specific transcriptional regulatory pathways.

INTRODUCTION

Genome-wide association studies (GWAS) have successfully identified over 10,000 common SNPs associated with hundreds of human traits and diseases (Welter et al., 2014). Each GWAS "hit" usually represents, or tags, hundreds of variants that are inherited together across a large (many are up to ~0.5 megabase) genomic region, termed a linkage disequilibrium (LD) block, often containing numerous protein-coding genes (Raychaudhuri, 2011). It is estimated that ~80% of the phenotypic heritability in common diseases and traits can be explained by non-coding regulatory variants (85%–90% of GWAS hits tag only non-coding variants) (Gusev et al., 2014), making target-gene identification and subsequent biological inference a considerable challenge (Edwards et al., 2013; Welter et al., 2014). However, these GWAS-nominated variants are significantly enriched at cell-type-specific regulatory regions such as DNase I hypersensitivity sites (DHS) and transcription factor (TF) occupancy sites, suggesting the attractive hypothesis that many of these variants may alter the regulation of gene transcription (Kundaje et al., 2015; Schaub et al., 2012).

Firmly establishing the causality of a GWAS-nominated regulatory variant requires clearly demonstrating its molecular functionality, identifying its target gene(s), and proving a connection to the original phenotype. Typically, identifying putative causal regulatory variants from GWAS requires a combination of genetic fine mapping, epigenomic profiling, and individual reporter assays (Edwards et al., 2013). Moving from putative causal variant (PCV) to target gene is facilitated either by expression quantitative trait loci (eQTL) studies in appropriate tissues or by creating isogenic cellular models (e.g., via genome editing) to identify the target gene(s). A target gene is then modulated in vitro in primary cell culture or in vivo in animal models to identify its role in determining the original phenotype. In a small number of cases, systematic approaches have successfully identified PCV(s), their mechanisms of action, target gene(s), and biological relevance for individual GWAS hits (Bauer et al., 2013; Claussnitzer et al., 2015; Edwards et al., 2013; Musunuru et al., 2010; Sankaran et al., 2008, 2012b).

In order to better understand the underlying biology behind an exponentially growing number of genetic associations, the development of scalable and high-throughput approaches is necessary. A recent study investigated loci associated with several autoimmune disorders by integrating finely mapped genetic associations from over 25,000 individuals with extensive enhancer annotations for 56 potentially relevant cell types, identifying a large number of PCVs (Farh et al., 2015). Nevertheless, this method resulted in identification of a single PCV for only ~10% of genetic associations. Other creative approaches acknowledge the expense and difficulty of genetic and epigenetic fine mapping and have leveraged phylogenetic information