Cell

From Structure to Systems: High-Resolution, Quantitative Genetic Analysis of RNA Polymerase II

Hannes Braberg,[1,4] Huiyan Jin,[5] Erica A. Moehle,[2] Yujia A. Chan,[6] Shuyi Wang,[1,4] Michael Shales,[1,4] Joris J. Benschop,[7] John H. Morris,[3] Chenxi Qiu,[5] Fuqu Hu,[5] Leung K. Tang,[5] James S. Fraser,[1,4] Frank C.P. Holstege,[7] Philip Hieter,[6] Christine Guthrie,[2,*] Craig D. Kaplan,[5,*] and Nevan J. Krogan[1,4,8,*]

[1]Department of Cellular and Molecular Pharmacology
[2]Department of Biochemistry and Biophysics
[3]Resource on Biocomputing, Visualization, and Informatics
University of California, San Francisco, San Francisco, CA 94158, USA
[4]California Institute for Quantitative Biosciences, QB3, San Francisco, CA 94158, USA
[5]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA
[6]Departments of Medical Genetics and Biochemistry, Michael Smith Laboratories, University of British Columbia, Vancouver V6T 1Z4, Canada
[7]Molecular Cancer Research, University Medical Centre Utrecht, 3508 AB Utrecht, the Netherlands
[8]J. David Gladstone Institutes, San Francisco, CA, 94158 USA
*Correspondence: christineguthrie@gmail.com (C.G.), cdkaplan@tamu.edu (C.D.K.), nevan.krogan@ucsf.edu (N.J.K.)
http://dx.doi.org/10.1016/j.cell.2013.07.033

SUMMARY

RNA polymerase II (RNAPII) lies at the core of dynamic control of gene expression. Using 53 RNAPII point mutants, we generated a *p*oint mutant *e*pistatic *m*iniarray *p*rofile (pE-MAP) comprising ~60,000 quantitative genetic interactions in *Saccharomyces cerevisiae*. This analysis enabled functional assignment of RNAPII subdomains and uncovered connections between individual regions and other protein complexes. Using splicing microarrays and mutants that alter elongation rates in vitro, we found an inverse relationship between RNAPII speed and in vivo splicing efficiency. Furthermore, the pE-MAP classified fast and slow mutants that favor upstream and downstream start site selection, respectively. The striking coordination of polymerization rate with transcription initiation and splicing suggests that transcription rate is tuned to regulate multiple gene expression steps. The pE-MAP approach provides a powerful strategy to understand other multifunctional machines at amino acid resolution.

INTRODUCTION

Alterations within a genome often cause specific as well as global phenotypic changes to a cell. Combining two alterations in the same cell allows for measurement of the genetic interaction between them: negative genetic interactions (synthetic sick/lethal) arise when two mutations in combination cause a stronger growth defect than expected from the single mutations. This is often observed for factors participating in redundant pathways or as nonessential components of the same essential complex. In contrast, positive interactions occur when the double mutant is either no sicker (epistatic) or healthier (suppressive) than the sickest single mutant (Beltrao et al., 2010) and may indicate that the factors are components of a nonessential complex and/or that the factors perform antagonizing roles in the cell. However, single genetic interactions are often difficult to interpret in isolation; an interaction pattern for a given mutation can be more informative, as it reports on the phenotype in a large number of mutant backgrounds (Schuldiner et al., 2005; Tong et al., 2004). These genetic profiles provide highly specific readouts that can be used to identify genes that are functionally related (Beltrao et al., 2010).

One of the first organisms to be genetically interrogated on a large scale was *Saccharomyces cerevisiae* (*S. cerevisiae*), in which nonquantitative genetic interaction data could be collected using the SGA (synthetic genetic array) (Tong et al., 2004) or dSLAM (heterozygous diploid-based synthetic lethality analysis on microarrays) (Pan et al., 2004) approaches. We developed a technique termed epistatic miniarray profile (E-MAP) (Collins et al., 2010; Schuldiner et al., 2005; Schuldiner et al., 2006), which utilizes the SGA methodology and allows for the quantitative collection of genetic interaction data on functionally related subsets of genes, including those involved in chromatin regulation (Collins et al., 2007b), RNA processing (Wilmes et al., 2008), signaling (Fiedler et al., 2009), or plasma membrane function (Aguilar et al., 2010). However, the vast majority of systematic genetic screening interrogates deletions of nonessential genes or hypomorphic knockdown alleles of essential genes. Because many genes, especially essential ones, are multifunctional, these methods perturb all activities associated with a given gene product.

Here, we describe an important advance of the E-MAP approach, which allows us to address higher levels of complexity

by examining the genetic interaction space of point mutant alleles of multifunctional genes in a technique that we term point mutant E-MAP (pE-MAP). This method greatly increases the resolution achievable by gene function analysis, as it allows assignment of specific genetic relationships to individual residues and domains. In this study, we have used the pE-MAP approach to functionally dissect RNAPII using alteration-of-function alleles in five different subunits of the enzyme. Using the genetic data, we assign transcriptional activity and specific functions to different residues and regions of RNAPII. By examining the relationship between transcription rate and genetic interaction partners, transcription-rate-sensitive factors were revealed. Through the characterization of multiple stages of gene regulation, including start site selection, transcription elongation rate, and mRNA splicing, the pE-MAP technique has provided both global and specific insight into structure-function relationships of RNAPII. We propose this strategy as a useful paradigm for the high-resolution interrogation of any multifunctional protein.

RESULTS

A Set of Alleles for the Functional Dissection of RNAPII

To identify residues that are important for transcriptional regulation in vivo, we isolated RNAPII alleles that confer one or more of the following transcription-related phenotypes: suppression of galactose sensitivity in *gal10Δ56* (Gal^R) (Greger and Proudfoot, 1998; Kaplan et al., 2005), the Spt^- phenotype (Winston and Sudarsanam, 1998), or mycophenolic acid (MPA) sensitivity (Shaw and Reines, 2000) (for additional details, see Figure 1A and the Experimental Procedures). Each of these phenotypes relates to a gene-specific transcription defect that can be monitored using plate assays (Figure 1A). Random mutagenesis by PCR was carried out on the entire coding regions of RNAPII subunit genes *RPB2*, *RPB3*, *RPB7*, and *RPB11* and most of *RPO21/RPB1* (Experimental Procedures). These genes encode the essential subunits that are unique to RNAPII (Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12 are shared with RNAPI and RNAPIII, and Rpb4 and Rpb9 are nonessential) (Archambault and Friesen, 1993). In total, 53 single point mutants were identified that exhibit at least one of these phenotypes (Kaplan et al., 2012) (Figure 1B and Figure S1 and Table S1 available online).

Analysis of the distribution of phenotypes relative to the RNAPII structure suggested that our alleles might be diverse in their functions. Gal^R and MPA-sensitive mutations were broadly distributed, whereas those with the Spt^- phenotype were less common and more localized (Figure 1B and Table S1). The screens identified mutations in highly conserved residues and structural domains known to be important for RNAPII activity, including the Rpb1 trigger loop, the Rpb1 bridge helix, and the Rpb2 lobe and protrusion (Cramer et al., 2001; Gnatt et al., 2001; Kaplan, 2013) (Table S1). Quantitatively measuring the genetic interactions of specific residues might provide insight into the functions of these RNAPII regions and could therefore also identify protein-protein interaction interfaces.

An RNAPII Point Mutant Epistatic Miniarray Profile

The 53 RNAPII point mutants (Figure 1B) were crossed against ~1,200 deletion and DAmP (decreased abundance by mRNA perturbation) alleles (Schuldiner et al., 2005) (Table S1), which represent all major biological processes. Thus, a quantitative pE-MAP comprising ~60,000 double mutants was created (Table S2; http://interactome-cmp.ucsf.edu). Two-dimensional hierarchical clustering of these data effectively grouped together genes from known complexes and pathways based on their interactions with the point mutants (Figure S2 and Data S1). Previous studies have demonstrated that genes encoding proteins that physically associate often have similar genetic interaction profiles (Collins et al., 2007b; Roguev et al., 2008). The data derived from the point mutants could differ in this respect, as it is based on only five subunits of a single molecular machine. A receiver operating characteristic (ROC) curve was therefore generated to measure how well the genetic profiles of the deletion and DAmP mutants in the pE-MAP predict known physical interactions between their encoded proteins (Experimental Procedures). It was found that the predictive power of the pE-MAP is similar to that of a previously published E-MAP (Collins et al., 2007b), indicating that the genetic interactions of the RNAPII point mutants report on connections among virtually all cellular processes (Figures 2A and S2 and Data S1).

Next, to examine whether the spatial location of a mutated residue is a determinant of its function, we compared the similarity of pairs of RNAPII genetic profiles to the three-dimensional distance between the mutated residues (Wang et al., 2006). We observed a strong correlation ($r = -0.37$, $p < 10^{-22}$) (Table S2), and the trend is significant both for residue pairs residing in the same subunit ($r = -0.25$, $p < 10^{-5}$) and for those in different subunits ($r = -0.28$, $p < 10^{-6}$) (Figure 2B). This suggests that structural proximity correlates with functional similarity and that high-resolution genetic interaction profiling could provide information for targets whose structures have not yet been determined.

Comparison of Genetic and Gene Expression Profiles Derived from the RNAPII Alleles

To determine whether any given genetic interaction might result from the point mutation affecting the expression of the corresponding gene, we subjected 26 of the RNAPII mutants to genome-wide gene expression analysis (Table S2; GEO accession number: GSE47429). We found no correlation ($r = -0.003$) between an RNAPII mutant's genetic interaction score with a gene deletion or DAmP allele and the expression change of that gene due to the RNAPII mutation (Figure 2C). Therefore, connections must be due to more complex relationships between the mutated residues and the library genes. Nonetheless, these data sets allowed us to test whether the clustering of the RNAPII mutants in the pE-MAP (Figure S2 and Data S1) could be recapitulated using their gene expression profiles. We thus assessed pair-wise RNAPII mutant similarity based on genetic and gene expression profiles separately and found that these two measures are highly correlated ($r = 0.71$; Figure 2D). Therefore, these orthogonal data sets provide a common biological framework for functionally organizing the RNAPII mutants, allowing us to study the underlying biology behind these mutants and their phenotypes.