LETTERS

Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map

Sean R. Collins[1,2,3], Kyle M. Miller[4], Nancy L. Maas[4], Assen Roguev[1,2], Jeffrey Fillingham[5], Clement S. Chu[1,2,3], Maya Schuldiner[1,2,3], Marinella Gebbia[5], Judith Recht[6], Michael Shales[5], Huiming Ding[5], Hong Xu[5], Junhong Han[7], Kristin Ingvarsdottir[8], Benjamin Cheng[9], Brenda Andrews[5], Charles Boone[5], Shelley L. Berger[8], Phil Hieter[9], Zhiguo Zhang[7], Grant W. Brown[10], C. James Ingles[5], Andrew Emili[5], C. David Allis[6], David P. Toczyski[4], Jonathan S. Weissman[1,2,3], Jack F. Greenblatt[5] & Nevan J. Krogan[1,2]

Defining the functional relationships between proteins is critical for understanding virtually all aspects of cell biology. Large-scale identification of protein complexes has provided one important step towards this goal; however, even knowledge of the stoichiometry, affinity and lifetime of every protein–protein interaction would not reveal the functional relationships between and within such complexes. Genetic interactions can provide functional information that is largely invisible to protein–protein interaction data sets. Here we present an epistatic miniarray profile (E-MAP)[1] consisting of quantitative pairwise measurements of the genetic interactions between 743 *Saccharomyces cerevisiae* genes involved in various aspects of chromosome biology (including DNA replication/repair, chromatid segregation and transcriptional regulation). This E-MAP reveals that physical interactions fall into two well-represented classes distinguished by whether or not the individual proteins act coherently to carry out a common function. Thus, genetic interaction data make it possible to dissect functionally multi-protein complexes, including Mediator, and to organize distinct protein complexes into pathways. In one pathway defined here, we show that Rtt109 is the founding member of a novel class of histone acetyltransferases responsible for Asf1-dependent acetylation of histone H3 on lysine 56. This modification, in turn, enables a ubiquitin ligase complex containing the cullin Rtt101 to ensure genomic integrity during DNA replication.
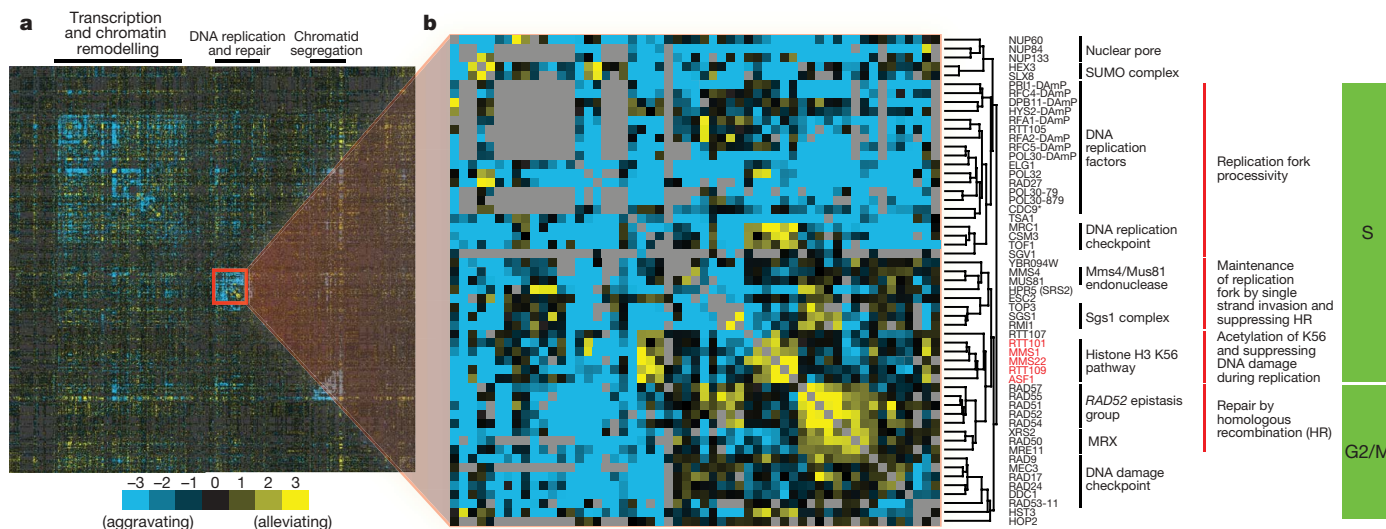
The synthetic genetic array (SGA)[2] and diploid-based synthetic lethality analysis on microarray (dSLAM)[3] approaches have enabled systematic identification of synthetic sickness/lethal (SSL) relationships in *S. cerevisiae* in which pairs of gene deletions are far more deleterious together than either of the individual deletions. Although individual SSL interactions can be difficult to interpret, the patterns of genetic interactions for gene mutations can be more informative because they provide high-resolution phenotypes that can be compared to identify functionally related genes[1–5]. Recently, we exploited the SGA strategy for generating double mutants to develop an approach, termed E-MAP[1], that greatly facilitates such comparisons. An E-MAP comprises comprehensive and quantitative measurements of genetic interactions between pairs of mutations within a defined subset of genes linked to one or more specific biological

processes[1]. E-MAPs are created by systematically generating yeast strains carrying each pair of mutations and measuring their growth rates. Genetic interactions are determined by comparing the observed fitness of the double mutants to an empirically determined typical fitness that would be expected on the basis of the growth defects associated with each mutation[1,6]. This technique allows for the identification of not only negative (aggravating) interactions, such as SSL pairs, but also positive (alleviating) interactions. Positive interactions include suppression, in which double mutants are healthier than the sicker of the two single mutants, as well as cases in which loss of one gene masks the effect of losing another, as is seen when two proteins act together in a common complex or pathway.

We comprehensively evaluated pairwise genetic interactions for 754 alleles of 743 genes involved in various aspects of chromosome biology (Fig. 1a; see also Supplementary Fig. 1 and Supplementary Data). The mutations include deletions of 663 non-essential genes and constitutive hypomorphic alleles—constructed using the 'decreased abundance by messenger RNA perturbation' (DAmP) strategy[1]—for 70 essential genes. Genes were selected based on published functional studies, protein–protein interaction data[7,8], earlier genome-wide SSL studies[2] and chemical sensitivity screens.

The resulting E-MAP consists of a 754 by 754 matrix of genetic interaction scores, where each row corresponds to the pattern (or profile) of interactions for one mutant allele of a gene (Fig. 1a). Using hierarchical clustering, we reordered the matrix to sort genes according to the similarity of their genetic interaction profiles. The resulting map has a modular structure that distinguishes between major processes such as transcription and chromatin remodelling, DNA replication and repair, and sister chromatid segregation. We illustrate the high-resolution functional information within these modules by focusing on a subcluster containing genes involved in DNA replication and repair (Fig. 1b). The general DNA replication factors (for example, RPA (*RFA1* and *RFA2*) and RFC processivity clamp loader subunits (*RFC4* and *RFC5*)) cluster near each other, and are resolved from the DNA replication checkpoint complex, Mrc1–Csm3–Tof1. The E-MAP also distinguishes groups of genes involved in sensing and repairing DNA damage including the *RAD52* epistasis group (*RAD51, RAD52, RAD54, RAD55, RAD57*), the MRX complex

Figure 1 | Hierarchical clustering of genetic interaction patterns. a, Full 'clustergram' of the patterns of interactions for all 754 mutations. Black horizontal bars indicate regions of the cluster corresponding to sets of genes implicated in the indicated functional processes. **b,** An enlargement of the 'DNA replication and repair' subcluster from **a**. The dendrograms indicate the relative similarities of the full patterns of interactions for the indicated genes. Various subclusters are annotated according to their known functions. Blue and yellow represent negative and positive genetic interactions, respectively. Grey boxes correspond to missing data points. Here, and throughout, genes indicated with an asterisk correspond to deletions of spurious ORFs that overlap the indicated gene. Genes highlighted in red represent novel findings that are referred to in the text.

(*RAD50*, *MRE11*, *XRS2*) and the 9-1-1 clamp (*DDC1*, *MEC3*, *RAD17*). The complete genetic interaction map, a useful resource for future functional studies, is available in Supplementary Data.
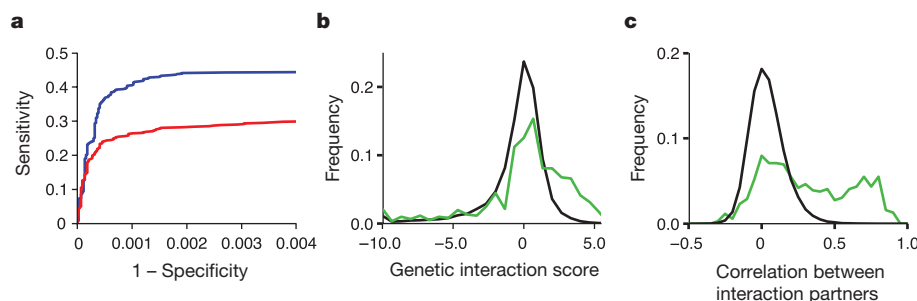
Beyond allowing simple hierarchical clustering, patterns of genetic interactions provide an unbiased way to identify sets of genes that function together in a coherent manner[1–3,5]. If two proteins act together to carry out a common function, one would expect deletions of the two encoding genes to have highly similar profiles of genetic interactions, as deletion of either gene would disrupt the same cellular process. Similarly, one would expect a positive genetic interaction between the two deletions, because in the context of the first, the second deletion would incur no additional cost. This relationship can be formalized in a COP (complex or pathway) score (see Supplementary Methods)[1,6], which synthesizes both expectations to create a single mathematical metric. Sets of genes connected by high COP scores are analogous to classically defined epistasis groups such as the well-studied *RAD52* epistasis group (Fig. 1b and Supplementary Fig. 2).

Recently, we used large-scale affinity purification data[7,8] to generate a physical interaction map that quantitatively reports through a purification enrichment (PE) score on the relative likelihood of each protein–protein interaction (see http://interactome-cmp.ucsf.edu)[9].

The accuracy and completeness of this integrated physical interaction map and the present E-MAP now make it possible to explore broadly the relationship between physical complexes and genetically defined epistasis groups. To evaluate the predictive power of the COP score relative to the physically based PE score, we used the protein complexes in the *Saccharomyces* Genome Database (SGD)[10] to define a trusted reference set of 'true positives' and 'true negatives' (see Supplementary Methods). We then used receiver operating characteristic (ROC) curves[11], which measure the rate at which each approach identifies true positives versus true negatives, to compare the predictive power of the two approaches.

Notably, the COP score identifies a distinct and large subset of protein–protein interactions with a specificity rivalling that of affinity purification (Fig. 2a; see also Supplementary Fig. 2). A key value of the E-MAP, therefore, is that it divides physical interactions into two classes: one group in which the proteins function coherently and a second in which their patterns of genetic interactions indicate that the proteins carry out distinct or even opposing functions. In particular, for pairs of physically interacting proteins, the histogram of either the genetic interaction scores or the correlation between genetic interaction patterns shows a roughly bimodal character (Fig. 2b, c). Thus, for a large fraction (somewhat greater than half) of physical

Figure 2 | Relationship between genetic epistasis groups and physical complexes. a, ROC curves comparing the power of the genetic interaction patterns—using the COP score (red) (see Supplementary Methods)—and large-scale affinity purification data (blue)—using a recent re-analysis of raw purification data[9]—to predict co-membership of pairs of proteins in the same physical complex. The slope of the initial portion of each curve serves as a measure of the score's maximal accuracy. **b,** Distribution of direct genetic interaction scores for pairs of genes encoding physically letteracting proteins (green) and non-interacting proteins (black) (see Supplementary Data). **c,** Distribution of the Pearson's correlation coefficients between the interaction patterns for the same sets of gene pairs as in **b**.
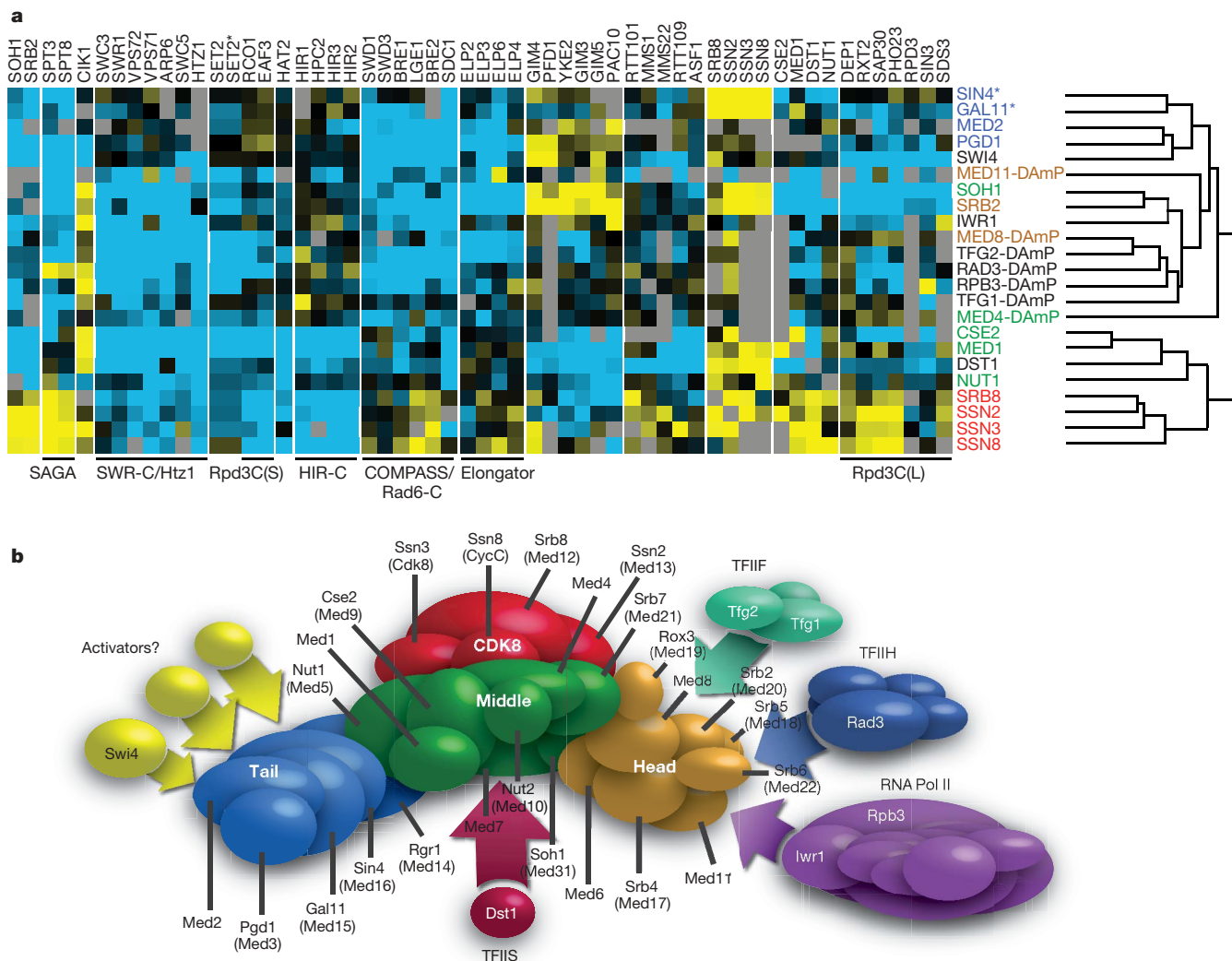
interactions, individual deletions of the two genes cause largely un-related phenotypes (see Supplementary Data for a complete comparison between genetic and physical complexes).

This ability of genetic interaction maps to reveal whether physically interacting proteins mediate a common function allows for the systematic dissection of multi-protein complexes into functional modules. We illustrate this by analysing Mediator, a conserved 25-protein complex critical for regulation of transcriptional initiation[12] by RNA polymerase II. Previous studies have suggested that Mediator is composed of four discrete physical modules: the head, middle, tail and CDK modules (Fig. 3b)[13,14]. Genetic interaction patterns accurately distinguish the protein components of these modules (Fig. 3a) and reveal distinct patterns of genetic interactions between components of different modules (Fig. 3a). For example, the CDK module displays positive genetic interactions with the head, middle and tail modules (Fig. 3a, b), consistent with the model in which the CDK module opposes the action of the remainder of the complex[14].

Our E-MAP data also reveal close functional relationships between Mediator components and non-Mediator proteins (Fig. 3a), thereby providing insights into specialized roles of the Mediator modules. For example, the transcriptional elongation factor TFIIS (Dst1) has a pattern of interactions highly correlated with patterns of middle module components, suggesting that this module may function with Dst1

during or after promoter escape. Conversely, the activator Swi4 clusters with the tail components, consistent with previous evidence that transcriptional activator proteins sometimes recruit Mediator through physical interactions with the tail module[13]. Finally, mutations of head module components behave most like mutations in the core transcriptional machinery (RNA polymerase II (*RPB3*, *IWR1*), TFIIF (*TFG1*, *TFG2*) and TFIIH (*RAD3*)), indicating that this module may have a critical role in the assembly and/or activity of the pre-initiation complex, which is consistent with a recent biochemical analysis[15].

In addition to dividing large protein complexes into functional sub-modules, the E-MAP-derived epistasis groups reveal many cases where proteins cooperate with each other even when they do not physically interact. One prominent and previously uncharacterized example of such an epistasis group is found within the DNA damage cluster (Fig. 1b) and comprises five genes: *RTT101*, *MMS1*, *MMS22*, *RTT109* and *ASF1* (Fig. 4a). Rtt101 is one of four cullins in budding yeast and is functionally[3] and physically associated with Mms1, Mms22, the E2 ubiquitin conjugating enzyme Cdc34, and the RING finger protein Hrt1 (refs 8–10, 16). This complex thus has features characteristic of ubiquitin ligases that target proteins for proteasomal degradation. *RTT109* is a poorly characterized open reading frame (ORF) identified as a regulator of Ty1 transposition[10]. Finally, Asf1 is a well-studied histone chaperone implicated in several

Figure 3 | Functional dissection of the Mediator complex. a, Subsets of the genetic interaction profiles for deletion or DAmP alleles of the indicated Mediator components (red, blue, green or orange font corresponding to the CDK, tail, middle and head modules, respectively) or non-Mediator genes (black type). The dendrogram indicates the relative similarities of the full
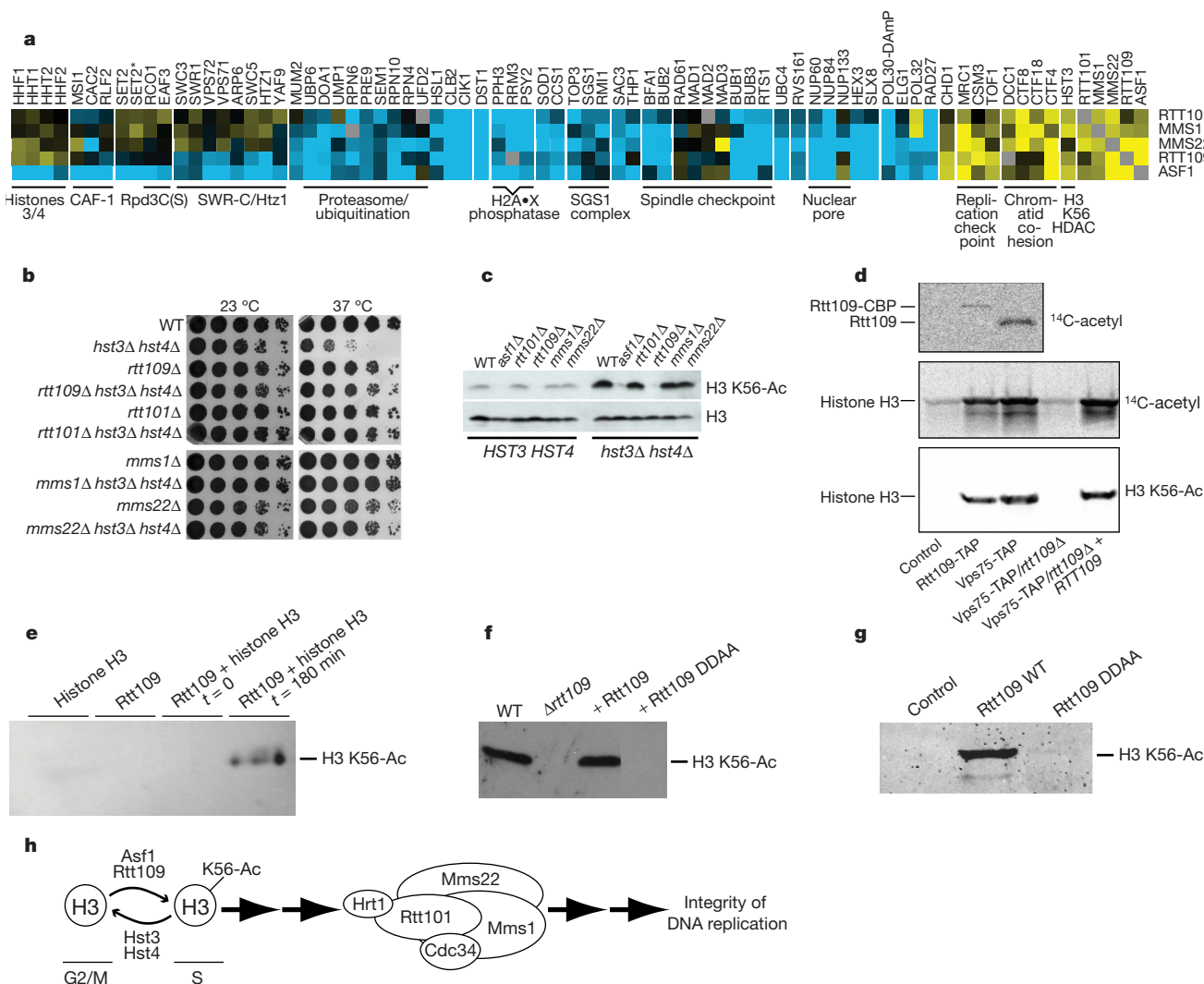
interaction patterns for the mutations shown. Blue and yellow represent negative and positive genetic interactions, respectively. Grey boxes correspond to missing data points. **b**, Modular organization of the Mediator complex[14] and close functional relationships between the Mediator modules and other transcriptional regulators.

processes including chromatin assembly[17], suppression of spurious transcriptional initiation[18] and the acetylation of histone H3 on lysine 56 (K56)[19].

Examination of interaction patterns suggests that the underlying functional connection between the members of this epistasis group is mediated through the cell-cycle-dependent acetylation of H3 K56 (refs 20–22). For example, deletion of *HST3*, which encodes the enzyme primarily responsible for H3 K56 deacetylation[21,22], shows positive genetic interactions with these genes (Fig. 4a). Consistent with this, deletion of *ASF1* (ref. 22), *RTT101*, *RTT109*, *MMS1* or *MMS22* suppresses the growth defect caused by hyperacetylation in *hst3Δ hst4Δ* strains (Fig. 4b). As seen for the *asf1Δ* strain, deletion of *RTT109* eliminates detectable K56 acetylation *in vivo* (Fig. 4c)[23]. Deletion or temperature-dependent inactivation of each known yeast histone acetyltransferase (HAT) protein has little effect on global levels of K56 acetylation *in vivo*[24], suggesting that Rtt109 itself might

be the responsible HAT. Indeed, we found that Rtt109, both affinity-purified from yeast (Fig. 4d) and recombinantly produced from *Escherichia coli* (Fig. 4e), acetylates itself as well as recombinant histone H3. Western blot analysis using a residue-specific antibody confirmed that Rtt109 acts on K56 (Fig. 4d). Mutation of two universally conserved adjacent aspartic acid residues abolished K56 acetylation both *in vivo* and *in vitro* (Fig. 4f, g). Notably, Rtt109 is itself acetylated on K290 (data not shown), just two amino acids away from the conserved aspartic acids, suggesting that auto-acetylation may be a mechanism of regulation. Finally, addition of Asf1 markedly enhances the *in vitro* H3 K56 acetylation activity of Rtt109 (Supplementary Fig. 4), suggesting that the Asf1 requirement for the acetylation *in vivo* reflects Asf1's ability to bind to and present H3–H4 heterodimers[25,26]. These observations indicate that Rtt109 is the founding member of a new family of HATs that shares no detectable sequence similarities with previously known HATs.

Figure 4 | Characterization of the histone H3 K56 acetylation pathway.
a, A subset of the genetic interaction patterns for *RTT101*, *MMS22*, *MMS1*, *RTT109* and *ASF1*. Shading as in Fig. 1. **b**, The growth of strains was assessed by plating (sevenfold) serial dilutions and incubating at either 23 °C or 37 °C. **c**, Western blot analysis was used to assess the K56 acetylation levels in wild-type, *asf1Δ*, *rtt101Δ*, *rtt109Δ*, *mms1Δ* and *mms22Δ* strains in *HST3 HST4* (wild type) or *hst3Δ hst4Δ* backgrounds. **d**, *In vitro* acetylation activity of Rtt109 purified from either Rtt109-TAP or Vps75-TAP strains (Rtt109 and Vps75 are complexed (Supplementary Fig. 3)[8]) was monitored by incorporation of radiolabel from ^{14}C-acetyl-coenzyme A (CoA) into either the purified Rtt109 or recombinant *Xenopus laevis* histone H3 purified from bacteria. No protein was added in the control reaction. Histone acetylation

was also monitored using antibody directed against K56 acetylated histone H3. CBP, calmodulin binding peptide. **e**, Recombinant histone H3, recombinant GST–Rtt109, or both, were incubated with acetyl-CoA. K56 acetylation was assessed by western blotting. **f**, Western blot analysis was used to assess K56 acetylation levels in a wild-type strain and in *rtt109Δ* strains episomally expressing either nothing (*Δrtt109*), wild-type Rtt109 (+ Rtt109), or Rtt109 mutated at D287A and D288A (+ Rtt109 DDAA). **g**, *In vitro* HAT assays were carried out using the indicated Rtt109 protein purified from Vps75-TAP strains. Results were analysed by western blot analysis. Mass spectrometry analysis confirmed that mutated Rtt109 (DDAA) still physically associates with Vps75 (data not shown). **h**, Model for the histone H3 K56 acetylation signalling pathway.

Unlike Rtt109 and Asf1, loss of Rtt101, Mms22 or Mms1 suppresses the *HST3*/*HST4* double deletion without preventing H3 K56 hyperacetylation (Fig. 4b, c). This suggests that the Rtt101 cullin complex is the major downstream effector of this acetylation pathway, thereby providing insight into this modification's physiological role. Specifically, defects in the progression of replication forks through damaged DNA and natural pause sites[27] seen in a *rtt101Δ* strain (as demonstrated by prolonged activation of the checkpoint protein Rad53 after methyl methanesulphonate treatment) are also seen when H3 K56 acetylation is blocked by deletion of *RTT109* or *ASF1* (Supplementary Fig. 3). Furthermore, *RTT101*, *MMS22* and *MMS1* deletions confer no additional sensitivity to DNA-damaging agents when combined with deletion of *ASF1* (J. Erkmann and P. Kaufman, personal communication). Taken together, these data suggest a model (Fig. 4h) in which H3 K56 acetylation acts upstream of the Rtt101 ubiquitin ligase complex—which is presumably targeting one or more proteins for degradation by the proteasome—to promote replication fork progression through damaged DNA. On the basis of positive genetic interactions in the E-MAP (Fig. 4a; see also Supplementary Data), potential targets include components of the replication checkpoint complex (Mrc1, Csm3, Tof1) and proteins involved in chromatid cohesion (Dcc1, Ctf4, Ctf8, Ctf18). Given that H3 K56 acetylation accumulates to ~50% during S phase[22], K56 acetylation may generally serve to mark newly synthesized DNA and allow the activity of Rtt101 to be specifically targeted to such regions during DNA synthesis. More broadly, this example shows that analysis of E-MAPs can illuminate complex biological pathways with high precision. Success here and in earlier work using hypomorphic alleles[1,28] for genetic interaction studies points to the potential for extending this approach to metazoans using technologies such as RNA interference.

METHODS

Strains were constructed and E-MAP experiments were performed as described previously[29]. Genetic interaction scores were computed as described in ref. 6. Histone acetyltransferase assays were performed as described[30], using immunoprecipitates from whole-cell extracts prepared from yeast strains containing one gene with a tandem affinity purification (TAP) tag[8]. More details for experimental assays, as well as a description of the data analysis, are provided in Supplementary Methods.

1. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
2. Tong, A. H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
3. Pan, X. *et al.* A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**, 1069–1081 (2006).
4. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotechnol.* **23**, 561–566 (2005).
5. Ye, P. *et al.* Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol. Syst. Biol.* **1**, 2005.0026 (2005).
6. Collins, S. R., Schuldiner, M., Krogan, N. J. & Weissman, J. S. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7**, R63 (2006).
7. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
8. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
9. Collins, S. R. *et al.* Towards a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* doi:10.1074/mcp.M600381-MCP200 (2 January 2007).
10. Hong, E. L. *et al. Saccharomyces* Genome Database ⟨http://www.yeastgenome.org/⟩ (7 March 2006).
11. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2001).
12. Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30**, 235–239 (2005).
13. Bjorklund, S. & Gustafsson, C. M. The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30**, 240–244 (2005).
14. van de Peppel, J. *et al.* Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* **19**, 511–522 (2005).
15. Takagi, Y. *et al.* Head module control of mediator interactions. *Mol. Cell* **23**, 355–364 (2006).
16. Michel, J. J., McCarville, J. F. & Xiong, Y. A role for *Saccharomyces cerevisiae* Cul8 ubiquitin ligase in proper anaphase progression. *J. Biol. Chem.* **278**, 22828–22837 (2003).
17. Tyler, J. K. Chromatin assembly. Cooperation between histone chaperones and ATP-dependent nucleosome remodeling machines. *Eur. J. Biochem.* **269**, 2268–2274 (2002).
18. Schwabish, M. A. & Struhl, K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell* **22**, 415–422 (2006).
19. Recht, J. *et al.* Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc. Natl Acad. Sci. USA* **103**, 6988–6993 (2006).
20. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436**, 294–298 (2005).
21. Maas, N. L., Miller, K. M., DeFazio, L. G. & Toczyski, D. P. Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol. Cell* **23**, 109–119 (2006).
22. Celic, I. *et al.* The sirtuins Hst3 and Hst4p preserve genome integrity by controlling histone H3 lysine 56 deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).
23. Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R. & Shilatifard, A. Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* **281**, 37270–37274 (2006).
24. Ozdemir, A. *et al.* Characterization of lysine 56 of histone H3 as an acetylation site in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 25949–25952 (2005).
25. Adkins, M. W., Carson, J. J., English, C. M., Ramey, C. J. & Tyler, J. K. The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase. *J. Biol. Chem.* **282**, 1334–1340 (2007).
26. English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. & Tyler, J. K. Structural basis for the histone chaperone activity of Asf1. *Cell* **127**, 495–508 (2006).
27. Luke, B. *et al.* The cullin Rtt101p promotes replication fork progression through damaged DNA and natural pause sites. *Curr. Biol.* **16**, 786–792 (2006).
28. Davierwala, A. P. *et al.* The synthetic genetic interaction spectrum of essential genes. *Nature Genet.* **37**, 1147–1152 (2005).
29. Schuldiner, M., Collins, S. R., Weissman, J. S. & Krogan, N. J. Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. *Methods* **40**, 344–352 (2006).
30. Mizzen, C. A., Brownell, J. E., Cook, R. G. & Allis, C. D. Histone acetyltransferases: preparation of substrates and assay procedures. *Methods Enzymol.* **304**, 675–696 (1999).