

# Widespread epistasis shapes RNA polymerase II active site function and evolution

Received: 2 April 2025

Accepted: 11 August 2025

Published online: 27 August 2025

 Check for updatesBingbing Duan<sup>1</sup>, Chenxi Qiu<sup>2,3</sup>, Sing-Hoi Sze<sup>4,5</sup> & Craig Kaplan<sup>1</sup>✉

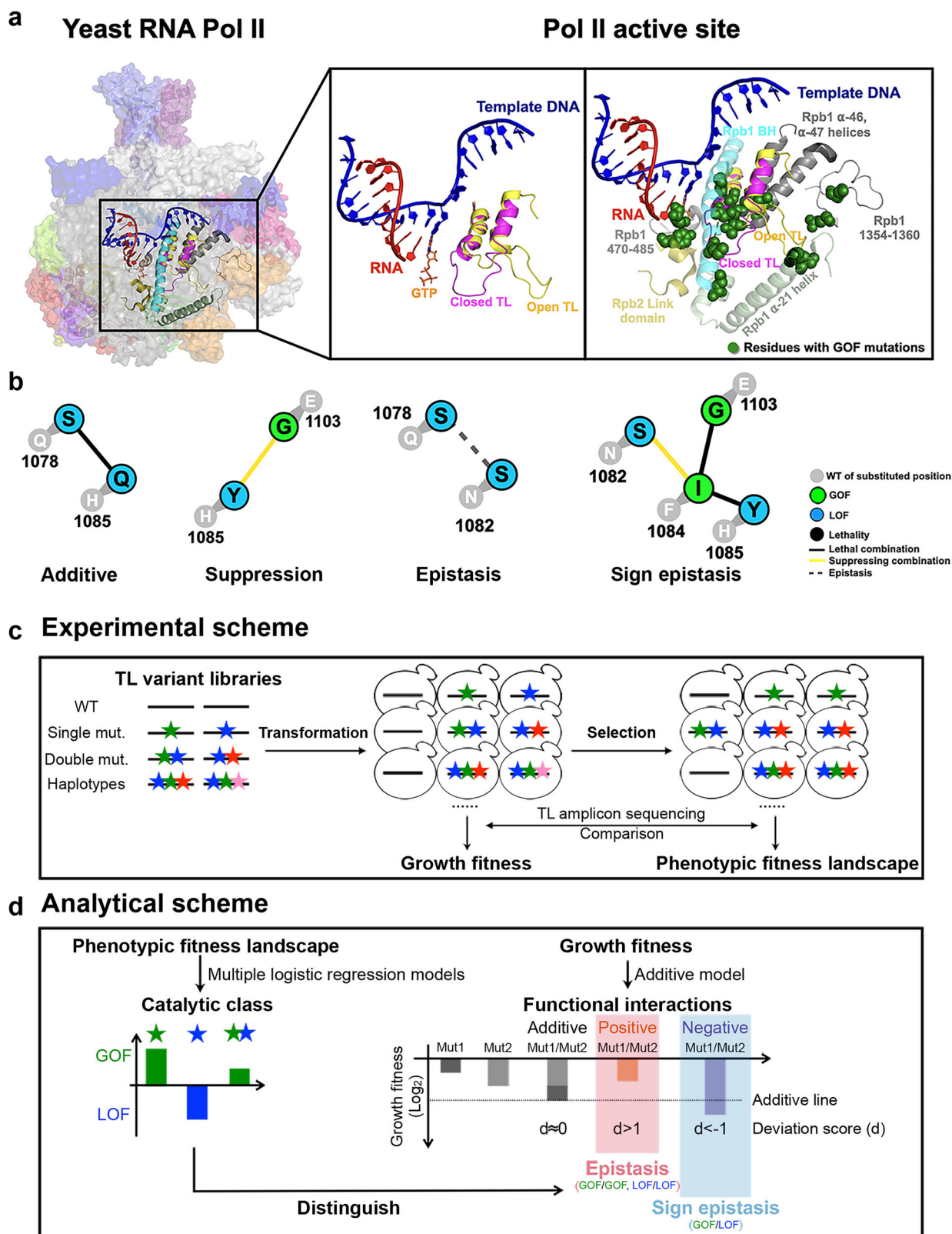
Multi-subunit RNA Polymerases are responsible for transcription in all kingdoms of life. These enzymes rely on dynamic, highly conserved active site domains such as the so-called “trigger loop” to accomplish steps in the transcription cycle. Mutations in the RNA polymerase II trigger loop confer a spectrum of biochemical and genetic phenotypes that suggest two main classes, which decrease or increase catalysis or other nucleotide addition cycle events. The RNA polymerase II active site relies on networks of residue interactions to function, and mutations likely perturb these networks in ways that may alter mechanisms. Here, we take a structural genetics approach to reveal residue interactions within and surrounding the RNA polymerase II trigger loop – determining its “interaction landscape” – by deep mutational scanning in *Saccharomyces cerevisiae* RNA polymerase II. This analysis reveals connections between trigger loop residues and surrounding domains, demonstrating that trigger loop function is tightly coupled to its specific enzyme context.

Transcription from cellular genomes is performed by conserved multi-subunit RNA polymerases (msRNAPs)<sup>1–3</sup>. While bacteria and archaea use a single msRNAP to transcribe all RNAs<sup>4–6</sup>, eukaryotes have at least three msRNAPs (Pol I, II, and III) for different RNA types<sup>7–10</sup>. RNA synthesis by msRNAPs occurs by iterative nucleotide addition cycles (NAC) of nucleotide selection, catalysis, and polymerase translocation<sup>11–15</sup>. The msRNAPs active sites accomplish these NAC steps using two conformationally flexible domains, the bridge helix (BH) and trigger loop (TL)<sup>5,12,16–19</sup>. The multi-functional natures of the BH and TL likely underlie their striking conservation, serving as interesting models for studying the function and evolution of extremely constrained protein domains.

Most catalytic cycle events involve concerted conformational changes in the TL, and potentially the BH<sup>12–14,19–23</sup>. The extent to which these conformations are supported by, or communicate with, the surrounding enzymatic context can be examined by structural or computational approaches, but functional importance needs to be determined experimentally. The BH is straight in most msRNAP

structures<sup>7,8,12,13,24</sup>, but appears kinked in *Thermus thermophilus* (bacteria) RNAP<sup>5,25</sup>. Simulations suggest these dynamics may promote msRNAP translocation<sup>16,25–28</sup>. The TL adopts various conformations to support its multiple functions. During each NAC, the TL nucleotide interaction region (NIR) discriminates correct NTPs from mismatched ones, initiating a shift from a catalytic-disfavoring “open” state to a catalytic-favoring “closed” state<sup>11,20,29,30</sup>. This closure promotes catalysis<sup>20,31</sup>, and subsequent TL reopening is proposed to support polymerase translocation to the next template DNA position, allowing for the next NAC<sup>32–34</sup> (Fig. 1a). Disrupting the TL with mutations affects transcription at all stages<sup>13,17,21,23,35–38</sup>. For instance, NIR mutations impair TL-substrate interactions, resulting in reduced catalysis and elongation rate in vitro (Loss of function, LOF)<sup>17,23,36,37,39,40</sup>. Mutations in the TL hinge and C-terminal regions appear to shift the TL toward the active state, enhancing catalysis and elongation rate but impairing transcription fidelity (Gain of function, GOF)<sup>17,41,42</sup>. Mutations may also affect other transcription processes such as proofreading, pausing, and termination<sup>12,42–45</sup>, as

<sup>1</sup>Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA. <sup>2</sup>Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. <sup>5</sup>Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX, USA. ✉e-mail: [craig.kaplan@pitt.edu](mailto:craig.kaplan@pitt.edu)



TL conformations are associated with these activities in addition to catalysis and translocation.

All TL conformational dynamics and functions are likely balanced by residue interactions within and around the TL<sup>40,44,46–48</sup>. Embedded in the conserved active site, the TL interacts with surrounding domains such as the BH and  $\alpha$ 46–47 helices, forming a five-helix bundle enclosing a hydrophobic pocket that supports these dynamics

(Fig. 1a)<sup>12,41</sup>. Mutations in TL adjacent domains (e.g., BH GOF T834P and LOF T834A, funnel helix  $\alpha$ -21 GOF S713P, and Rpb2 GOF Y769F)<sup>17,23,25,26,35–37</sup> (Fig. 1a), and in a Pol II subunit which does not directly contact the Pol II active site, such as Rpb9<sup>49</sup>, also alter catalytic activity and fidelity, showing phenotypes similar to TL mutants, indicating that TL dynamics and function are finely balanced and could be sensitive to allosteric effects<sup>49–52</sup>. Determining the extent and nature of

**Fig. 1 | Schematics of the Pol II active site interaction landscape.** **a** The Pol II active site is embedded in the center of a 12-subunit complex (left panel). Pol II functions are supported by distinct TL conformational states. An open TL (PDB: 5C4X) and closed TL (PDB: 2E2H) conformations are shown in the middle panel. GOF mutations have been identified in the TL and its proximal domains (right panel), suggesting TL mobility and function may be impacted by adjacent residues. **b** Examples of inter-residue genetic interactions. WT residues are shown in circles with a number indicating residue position in Rpb1. Mutant substitutions are shown in colored circles, with color representing mutant class. Colored lines between mutant substitutions represent types of genetic interactions. **c** Overview of experimental approach. We synthesized 10 libraries of TL variants represented by colored stars. Libraries were transformed into WT or mutated yeast strains. A

selection assay was subsequently performed by scraping and replating the transformants onto different media for phenotyping. DNA was extracted from yeast from all conditions, and then went through TL region amplification and Illumina sequencing. Read counts for variants on general conditions were used to determine growth fitness, while read counts on other conditions were used to determine the phenotypic fitness landscape. **d** Overview of analytical approach for determining interaction landscape. Mutant conditional growth fitnesses were calculated using allele frequencies under selective growth conditions and subjected to two logistic regression models for classification/prediction of catalytic defects. Double mutant interactions were computed using growth fitness. Classification allowed epistatic interactions to be deduced from double mutant growth fitness (see “Methods”).

the pathways connecting the TL to the rest of the polymerase will reveal the functional networks that shape its dynamics and the potential pathways through which Pol II has evolved or may be controlled by elongation factors.

Physical and functional intramolecular interactions between amino acids define protein function and evolvability<sup>53–56</sup>. Genetic interactions where one mutation modulates or alters the phenotype of another are commonly referred to as epistasis. Interactions can be of different types, such as suppressive, masking, or synthetic interactions. In this study, we use “epistasis” specifically to describe cases where mutant phenotypes allele-specifically depend on the identities of other amino acids. Epistasis contributes to protein evolvability by providing a physical context and an evolutionary window in which some intolerable mutations may be tolerated<sup>57–59</sup>. Recent studies have shown how mutations can alter protein function, allostery, and evolvability, suggesting that even conserved residues are subject to unique, context-dependent epistatic constraints<sup>60–67</sup>. Consistent with this, distinct phenotypes for the same conserved residues have been observed in various proteins, including yeast Pol I and Pol II<sup>65,66,68</sup>. For example, the TL domain of yeast Pol I is incompatible with Pol II, despite sharing 70% sequence identity, while the similarly distinct Pol III TL maintains some function<sup>69</sup>. It is not known generally if highly conserved TL sequences are compatible in a heterologous msRNAP context. Answering this question would reveal how tightly integrated the TL is by coevolution with its embedded context, or the extent to which its functions are self-contained due to its very high conservation.

Functional interactions between residues can be revealed by genetic interactions of double mutants<sup>17,37,70–72</sup>. A systematic analysis is needed to fully describe and understand the requirements for each mutant’s phenotype, and the nature of the interactions that control Pol II activity. Previous studies from our lab on a small subset of site-directed substitutions identified distinct types of Pol II double mutant interactions, including suppression, synthetic sickness and lethality (activity-additive), less-than-additive interactions (sometimes referred to as epistasis), and sign epistasis<sup>17,37,71</sup>. Suppression interactions are common between LOF and GOF mutants, as expected if each mutant is individually acting in the double mutant, balancing opposing effects on activity. Similarly, synthetic lethality or sickness were common between mutants of the same class (GOF or LOF), consistent with greater defects arising from combining partial losses of TL function. However as just noted, we also observed mutant combinations where the double mutant was not sicker than the single mutants for mutants of the same class, suggesting these mutants might act at the same step. Additionally, we had observed a specific type of epistasis called sign epistasis, where a mutant’s phenotypic class (GOF or LOF) appeared dependent on the identity of another residue. For example, the GOF TL substitution Rpb1 F1084I was unexpectedly lethal with the LOF TL substitution Rpb1 H1085Y, contrary to the predicted mutual suppression for a GOF/LOF combination<sup>37,71</sup>. This suggested that F1084I

required H1085 for its GOF characteristics and became a LOF mutant in the presence of H1085Y (Fig. 1b).

Deducing large-scale residue interaction networks is challenging. To accomplish this for Pol II, we developed experimental and analytical schemes to extend the previously established “Pol II phenotypic landscape”<sup>17,37,71,73</sup> to 15,174 double and multiple mutants within the *S. cerevisiae* Pol II TL and between the TL and adjacent domains (a “Pol II Interaction landscape”). Our analyses indicate that TL function and evolution are dominated by widespread epistasis. The TL faces extensive constraints despite its high conservation, arising from residues within the TL and its interacting domains, resulting in highly similar TLs being incompatible inside yeast Pol II to a large degree. Additionally, individual mutants with similar biochemical and genetic phenotypes have quite distinct genetic interaction networks, revealing differential functional requirements for their abilities to confer altered Pol II activities. Some mutants minimally change the genetic behavior of most TL substitutions when combined, suggesting that TL conformations or interactions remain largely unchanged, with altered TL dynamics hypothesized to be the sole difference. In contrast, other alleles greatly perturb the genetic behavior of many TL substitutions, indicating these substitutions change the active site landscape in ways that would otherwise be undetectable. Together, these results suggest small changes in the Pol II active site can dramatically change its constraints and therefore, its potential evolutionary paths. Furthermore, our findings highlight potential communication networks through evolutionary coupling of Pol II residues within the active site, which may underlie allosteric communication paths to the active site for elongation factors.

## Results

### A system for the dissection of the Pol II active site genetic interaction landscape

We developed an experimental and analytical framework, termed the Pol II TL interaction landscape, to dissect residue interactions that shape Pol II TL function and evolution in *S. cerevisiae*. We designed and synthesized 15,174 variants representing evolutionary haplotypes, all possible Pol II TL single mutants and a subset of targeted double mutants in ten libraries (Supplementary Table 1). This approach follows our prior analysis of the TL phenotypic landscape<sup>17</sup> with modifications (“Methods” and Supplementary Fig. 1a). Libraries were transformed, screened under selected conditions (Methods), and sequenced (Fig. 1c, Supplementary Fig. 1a). Mutant growth phenotypes are calculated based on relative allele frequency shifts compared to a control condition and normalized to the WT under identical conditions. Biological replicates indicate high reproducibility (Supplementary Fig. 1b, c). Libraries were Min-Max normalized<sup>74</sup> to account for scaling differences between libraries (Supplementary Fig. 2a), and the same mutants present among different libraries indicate high correlation of fitness determinations in each library (Supplementary Fig. 2b, c).

We defined a conceptual framework for evaluating genetic interactions among TL mutations. First, we assume that independent mutant effects combine in a log-additive manner. This means that predicted double mutant fitness defects should be the combination of both single mutant defects, as is standardly assumed<sup>59,70,72,75</sup>. Deviations from log-additive expectations represent genetic interactions. To quantify any potential deviation, we compared the observed fitness of a double mutant measured in our high-throughput experiment to its expected fitness, calculated as the addition of constituent single substitutions' fitnesses, which are expressed as a log value. We term these differences as a "deviation score" (Deviation score = observed fitness – expected fitness; Supplementary Fig. 3). We consider a deviation score between –1 and 1 to be consistent with additive behavior, meaning independent contributions of single mutants to the double mutant fitness. A deviation score > 1 is taken as a positive genetic interaction (e.g., suppression), while a score < –1 indicates a negative genetic interaction (e.g., synthetic sickness or lethality).

Importantly, our phenotyping analyses determine both general mutant fitness and predict their class of *in vitro* biochemical defect. Our previous studies showed that phenotypic profiles from conditional growth assays, reflecting Pol II-activity-dependent phenotypes *in vivo*, correlate with the measured biochemical activity *in vitro*<sup>17,37,76</sup>. For example, catalytically hyperactive GOF mutants have a phenotypic profile based on sensitivity to mycophenolic acid (MPA) and weak suppression of two genetic reporters (*gal10Δ56* and *lys2-128Δ*), while catalytically defective LOF mutants are MPA resistant, show strong suppression of *gal10Δ56*, and fail to suppress *lys2-128Δ* (more fully described in "Methods"). Using the established correlation between growth profiles (MPA and media used to detect phenotypes of mutants on *gal10Δ56* and *lys2-128Δ* backgrounds) and known/predicted *in vitro* catalytic effects (GOF or LOF) of 65 TL alleles<sup>17,37,52</sup>, we trained two multiple logistic regression models to distinguish Pol II mutant classes, and applied them to all viable mutants (growth fitness score > –6.5) of designed variants, classifying them into three groups, GOF (22.96%), LOF (28.11%) and unclassified (mutants not belonging to either class, 48.93%) (Fig. 2a, "Methods"). To visually inspect the classifications, we applied t-SNE projection and k-means clustering for all measured mutants across all growth conditions. We observed separated GOF and LOF clusters consistent with logistic regression classifications (Fig. 2b, c). Notably, mutants with similar defects may be further distinguished in different clusters, suggesting their similar phenotypic defects may be related to different transcription defects (Fig. 2c). By classifying Pol II active site mutants as "GOF" or "LOF" based on predicted *in vitro* biochemical phenotypes, we establish baseline expectations for their genetic interactions. We previously observed activity-additive interactions, meaning suppression between mutants of different classes (GOF + LOF) or synthetic sickness/lethality within the same class (GOF + GOF or LOF + LOF). Together with the calculated deviation scores, we can distinguish specific epistatic interactions from activity-dependent effects. We classified an interaction as epistasis when we observed positive fitness deviation in mutants of the same activity class (GOF + GOF, LOF + LOF), where we would otherwise expect synthetic sickness or lethality if mutants were functioning independently. Conversely, we defined sign epistasis for situations where negative deviations in combinations between the classes (GOF + LOF), where suppression would be expected for independent mutants (Fig. 1d, Supplementary Fig. 3).

We compared interactions observed in sequencing to growth assays for 59 previously examined double mutants<sup>17,37</sup>, with 48 matching our observations. The other 11 were categorized as ultra-sick/lethal and therefore inaccessible to sequencing analysis (Supplementary Table 2). We constructed 50 additional mutants to assess interactions by patch assay and spot assay, and these aligned with our high-throughput results (Supplementary Fig. 4).

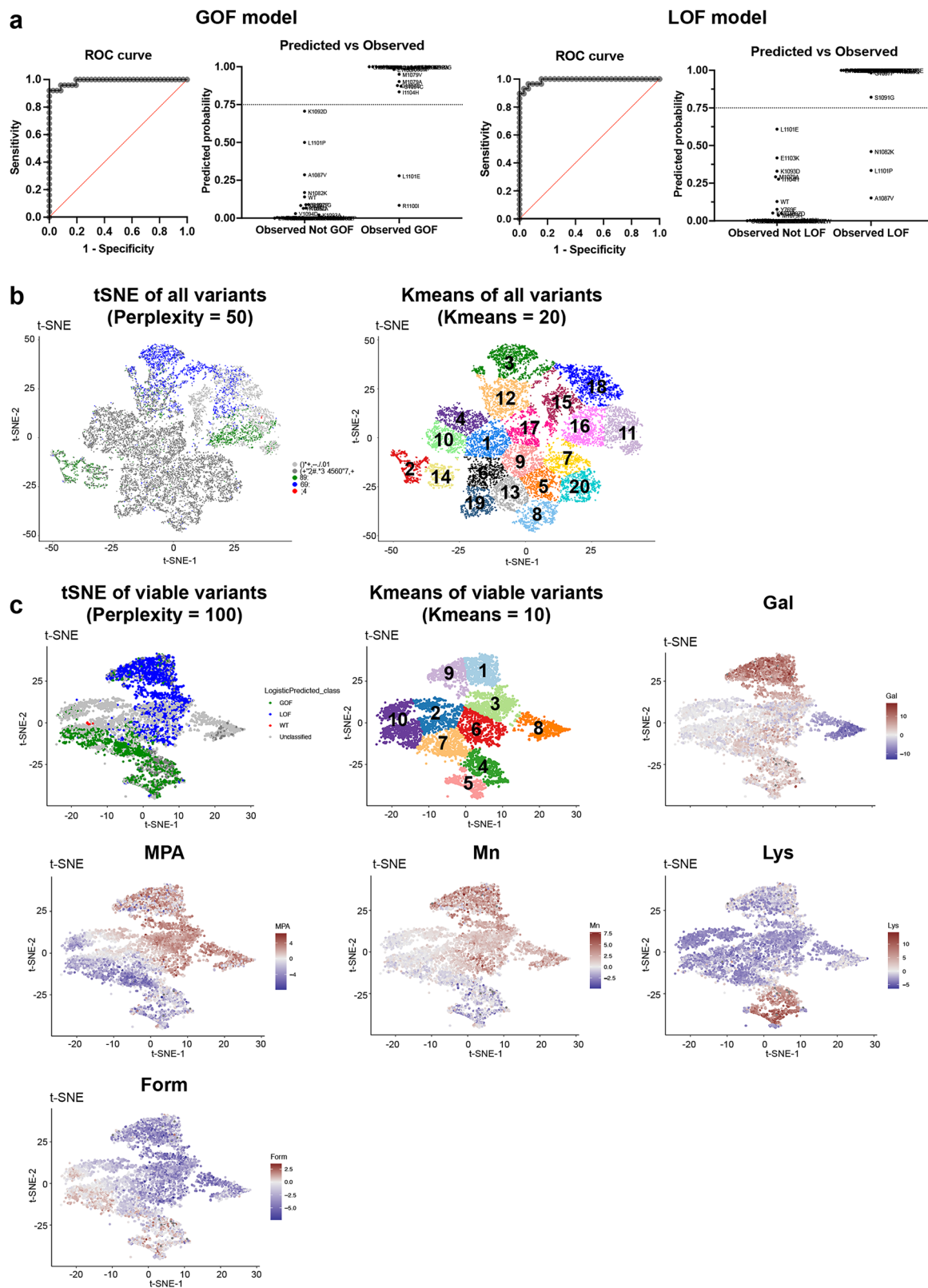
### Widespread incompatibility of conserved TL sequences with yeast Pol II suggests the TL is highly constrained by its coevolved environment

We previously found that identical mutations in a residue conserved between the Pol I and Pol II TLs yielded different biochemical phenotypes<sup>69,77</sup>. Additionally, the yeast Pol I TL was incompatible within the yeast Pol II enzyme, implying that TL function in that case depended on its enzymatic context<sup>69,77</sup>. To determine the scope of TL-Pol II incompatibility, we designed a library containing evolutionary TL variants from bacterial, archaeal, and eukaryotic msRNAPs and determined their compatibility in the yeast Pol II context (Fig. 3a). Eukaryotic TL alleles were more compatible than those from archaea or bacteria, with Pol II alleles being the most compatible (Fig. 3b, Supplementary Fig. 5a, b), consistent with evolutionary distance. Growth fitness in the Pol II background showed a slight negative correlation with the number of TL substitutions for most TLs but not those from bacteria (Supplementary Fig. 5c), likely because of the bacterial TLs being largely incompatible with Pol II. Despite this, some archaeal TLs provide yeast Pol II viability, yet several Pol II TLs were defective or lethal. These results suggest widespread coevolution between TL sequences and that the enzymatic background, beyond ultra-conserved positions, shapes TL function in individual enzymes. We have explored the nature of this incompatibility and complex intra-TL genetic interactions further in a separate study<sup>78</sup>.

### Allele-specific interactions suggest unique properties of individual mutants with similar phenotypes

To determine the potential residue interactions surrounding the TL in the yeast Pol II environment, we analyzed 12 previously studied GOF and LOF mutants (eight within the TL and four outside), each combined with >600 perturbations in the form of all single-substitution TL mutants (Fig. 4a). These 12 mutants serve as probes for the genetic interaction space of the TL, revealing how it might be altered in allele-specific fashion by each "probe" mutation. Importantly, TL conformational dynamics and function are balanced by internal residue interactions within the TL and by TL-external interactions between the TL and TL-proximal domains. While GOF and LOF mutants in TL-adjacent domains appear similar to GOF and LOF TL mutants, their underlying mechanisms or specific residue dependencies are not known. We therefore compared interaction networks among probe mutants with similar biochemical and phenotypic defects to distinguish if changes to TL function might reflect simple alterations to TL dynamics, or additional alterations to folding trajectories or conformations.

We observed distinct interaction networks among probe mutants, even for those with similar apparent biochemical and growth defects (Fig. 4b–d), indicating that mutants differentially impact the Pol II active site, potentially reflective of different constraints and distinct properties of individual mutants, as discussed below. A simple expectation is that mutants with altered TL dynamics yet maintenance of conformational states should show additive interactions, because the effects of substitutions on function would be maintained. In contrast, substitutions altering TL conformations/folding or propagating changes across the active site may be subject to allele-specific interactions, i.e., prominent epistasis. We observed that most probe mutants were placed on a continuum of effects based on the balance between additive (independent) interactions and epistatic (dependent or allele-specific) interactions in double mutants, which we interpret as the degree to which they propagate their effects across the active site (Fig. 4b, c, Supplementary Fig. 6, 7). Probe mutants appeared separable based on the extents of their additive or epistatic interactions, where greater numbers of independent interactions (additive) would indicate fewer effects on other residues' functions. In contrast, an increase in dependent interactions (sign epistasis) would indicate greater changes in other mutants' characters and presumably residue function. For

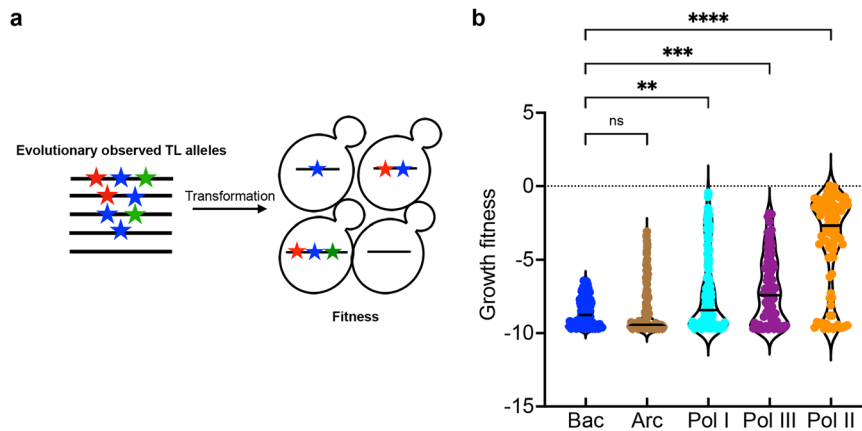


example, Y769F, a GOF TL-proximal mutant in the Rpb2 link domain (Fig. 1a), uniquely suppresses some TL lethal substitutions among all GOF probe mutants (Supplementary Fig. 8), indicating specific functional interactions between Y769F and these substitutions. Moreover, two TL-adjacent GOF mutants, Rpb1 S713P (funnel  $\alpha$ -helix 21) and T834P (BH) displayed largely distinct interaction networks despite similarly increased biochemical activity and genetic phenotypes as single

mutants<sup>17</sup> (Figs. 4c and 5a). S713P broadly suppressed LOF TL substitutions (96 instances), consistent with generic enhancement and preservation of TL dynamics or residue function. In contrast, T834P exhibited sparser suppression (33 instances) but much greater sign epistasis than S713P (102 versus 38 instances), especially with P1099, where all substitutions showed sign epistasis with T834P but not S713P (Fig. 5a). These results suggest that T834P's GOF function requires most

**Fig. 2 | Classification of mutant catalytic defects with machine learning algorithms.** **a** ROC curves of two multiple logistic regression models. Using 65 mutants with validated in vitro catalytic defects and conditional growth fitness, we trained two models to classify variants as GOF or LOF. The GOF AUROC is 0.9889 ( $P \leq 0.0001$ ), whereas the LOF ROC is 0.9914 ( $P \leq 0.0001$ ). The predicted vs. observed graphs show GOF/LOF probabilities for 65 known mutants. The threshold to determine GOF or LOF mutations is shown by lines at 0.75. Details are in Supplementary Table 6. Among the 6054 viable mutants, 1390 were classified as GOF (22.96%), 1702 as LOF (28.11%), and 2962 remained unclassified (48.93%). **b** Left: t-SNE projection of all mutants ( $n = 15174$ ) with perplexity = 50. Right: k-means clustering of all mutants. The t-SNE and k-means projections suggest GOF are in 3 clusters (cluster 2, 14, and 16), LOF are in 2 clusters (cluster 3 and 18), and unclassified mutants are in 2 clusters (11 and 15). Most ultra-sick/lethal mutants (fitness  $\leq -6.5$ ) are projected together into 13 clusters, likely due to significant noise

from low read counts across conditions. **c** Feature plot of viable mutations in t-SNE and k-means projections ( $n = 6054$ ). Ultra-sick/lethal mutations were removed, and the viable mutants were projected with t-SNE (perplexity = 100) and k-means (10 clusters). GOF were grouped into 4 clusters (4, 5, 7, and 10) and LOF were in 4 clusters (1, 3, 6, and 9). Each spot in the projection represents a mutant, and it is colored based on the fitness of the mutant in selective conditions. GOF and LOF mutants in different clusters are related to various phenotype patterns. GOF clusters 7 and 10 are defined by strong MPA<sup>S</sup>, while clusters 4 and 5 show slight MPA<sup>S</sup>, Gal<sup>R</sup>, Mn<sup>S</sup>, but strong Lys<sup>+</sup>. Slight Form<sup>S</sup> is a common feature across four GOF clusters. LOF clusters 3 and 6 show slight Mn<sup>R</sup>, while clusters 1 and 9 are strongly Mn<sup>R</sup> and Gal<sup>R</sup>. Cluster 8, which mostly contains unclassified mutants, appears defined by Gal super sensitivity, indicating a potential specific defect defining this cluster.



**Fig. 3 | Contextual epistasis shapes TL evolution.** **a** Schematic for the TL evolutionary haplotypes library. We selected 662 TL haplotypes representing TL alleles from bacterial, archaeal, and the three conserved eukaryotic msRNAPs. These TL alleles were transformed into yeast and were phenotyped under selective conditions. **b** Fitness of evolutionarily observed TL haplotypes in the yeast Pol II

background. The Pol II WT TL fitness (0) is labeled as a dotted line. Kruskal–Wallis test was performed for comparison, and significant levels (Bacteria vs. Archaea,  $P > 0.9999$ . Bacteria vs. Pol I,  $P = 0.0027$ . Bacteria vs. Pol III,  $P = 0.0003$ . Bacteria vs. Pol II,  $P < 0.0001$ ) were labeled.

TL residues to be WT, is especially reliant on proline at 1099, and that T834P alteration to the BH extensively changes the TL's functional space, likely through folding or altered conformations.

A similar distinction was observed between two internal TL GOF mutants, Rpb1 E1103G and F1084I (Figs. 4c and 5b). E1103G showed widespread suppression of LOF TL substitutions (184 instances), consistent with its proposed role in altering TL dynamics to promote TL closure<sup>23,37</sup>, and which allows TL mutants to maintain their effects. In contrast, F1084I showed more limited suppression (43 instances) while showing widespread synthetic lethality due to predicted sign epistasis, indicating that F1084I requires many WT residues for GOF characteristics and itself may switch to a LOF upon many additional TL perturbations.

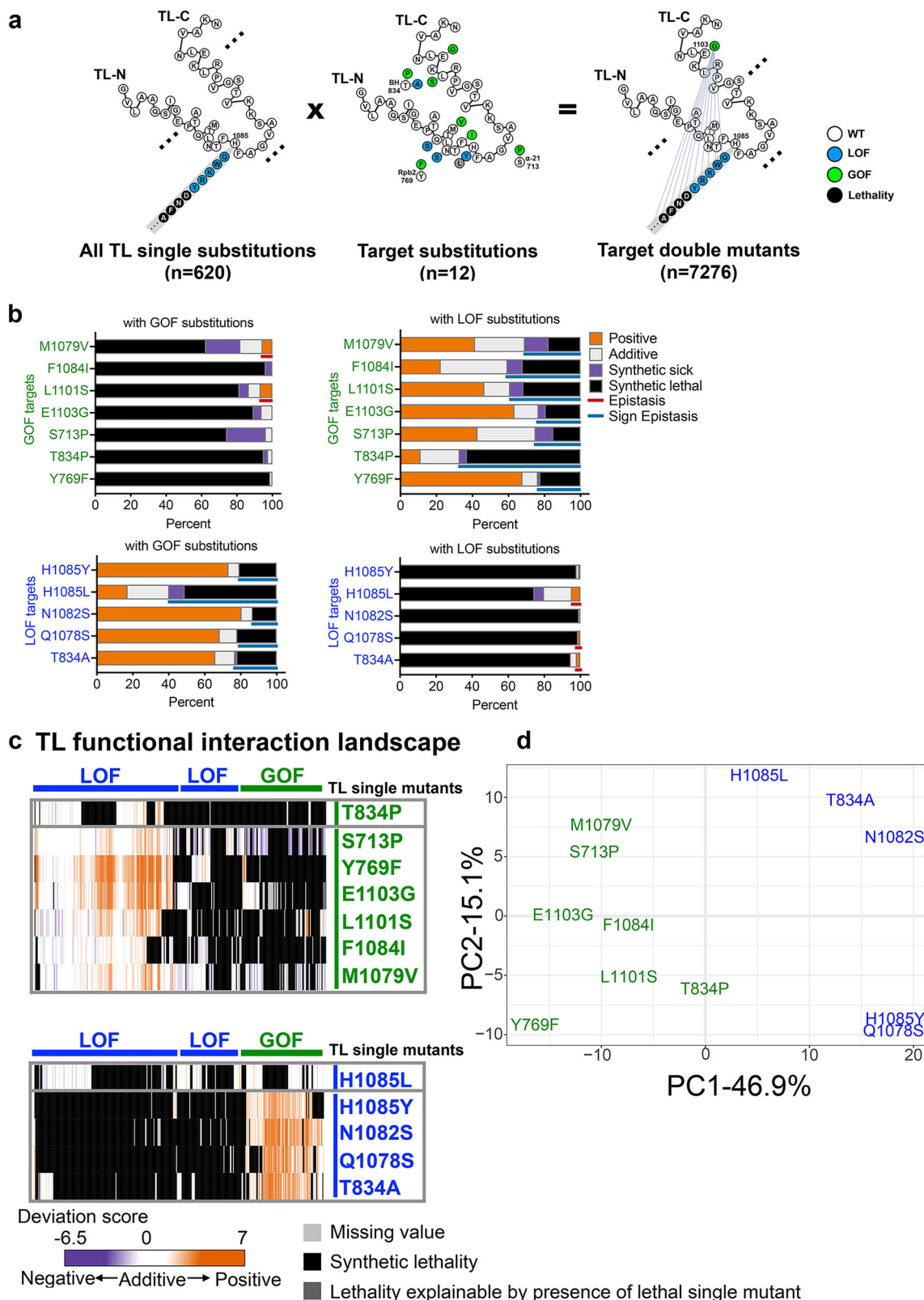
The most striking example came from two LOF substitutions at the ultra-conserved Rpb1 H1085 (Fig. 5c). This histidine, which contacts incoming NTP substrates<sup>5,12</sup>, is the target for the Pol II inhibitor  $\alpha$ -amanitin<sup>36,79</sup>, and promotes catalysis<sup>12,13</sup>. Initial structural data and molecular dynamics simulations suggested that H1085 functions as a general acid for Pol II catalysis<sup>80–83</sup>. Our discovery that H1085L was well-tolerated<sup>17</sup>, and subsequent experiments from the Landick lab<sup>84,85</sup>, have led to their proposal that the TL histidine functions as a positional catalyst where a similarly sized leucine substitution supports catalysis with mild effects on biochemistry and growth. If H1085Y and L substitutions represent a continuum of positional catalyst activity, then their interaction networks would be expected to be similar, and differing only in magnitude, not identity or type of interactions. Contrary to this, distinct interaction patterns were observed (Figs. 4c, and 5c). Most GOF mutants suppressed H1085Y but not H1085L. Instead,

H1085L showed synthetic lethality with most GOF mutants (putative sign epistasis). For example, almost all substitutions at E1103 showed sign epistasis with H1085L but not H1085Y (Supplementary Fig. 7b, Fig. 5c). The distinction between H1085L and H1085Y is further highlighted in PCA analysis (Fig. 4d).

### Functional and epistatic interactions within the TL hydrophobic pocket tune Pol II activity

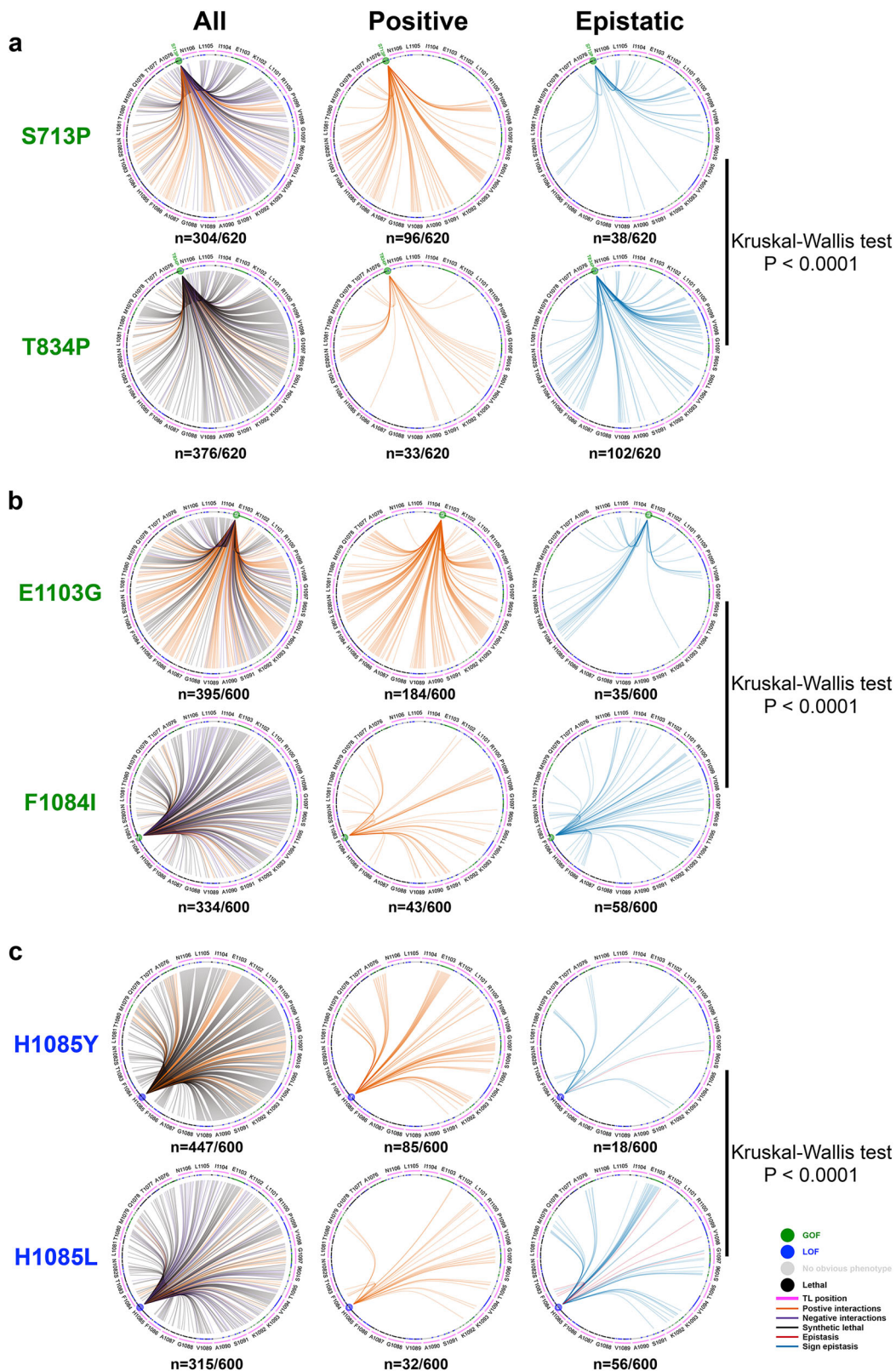
Several allele-specific epistatic interactions were observed, with some of the most extensive occurring between A1076 substitutions and L1101S (Fig. 6a), suggesting tight coupling between A1076 and L1101 for Pol II function. These two hydrophobic residues, together with other hydrophobic residues in TL helices, likely stabilize the open TL conformation through hydrophobic packing of the TL helices (Fig. 6c). Consistent with this, another pair of adjacent residues, M1079 and G1097, also showed allele-specific interactions (Fig. 6b, c). Additionally, three GOF/GOF combinations between L1101S and A1076 substitutions showed WT-like behavior for the double mutants based on our regression model for mutants, indicating mutual suppression due to one of the singles switching to a different class (sign epistasis), further supporting a tight functional interaction between 1101 and 1076 (Supplementary Fig. 9a).

We also observed allele-specific interactions for predicted lethal mutants. Our lethality threshold is more conservative than the actual lethality threshold, and therefore, some classified lethal but actually slow-growing mutants have enough growth in phenotypic assays for GOF or LOF class determination. Among 21 ultra-sick/lethal TL substitutions predicted as GOF, suppression by LOF mutants was expected and



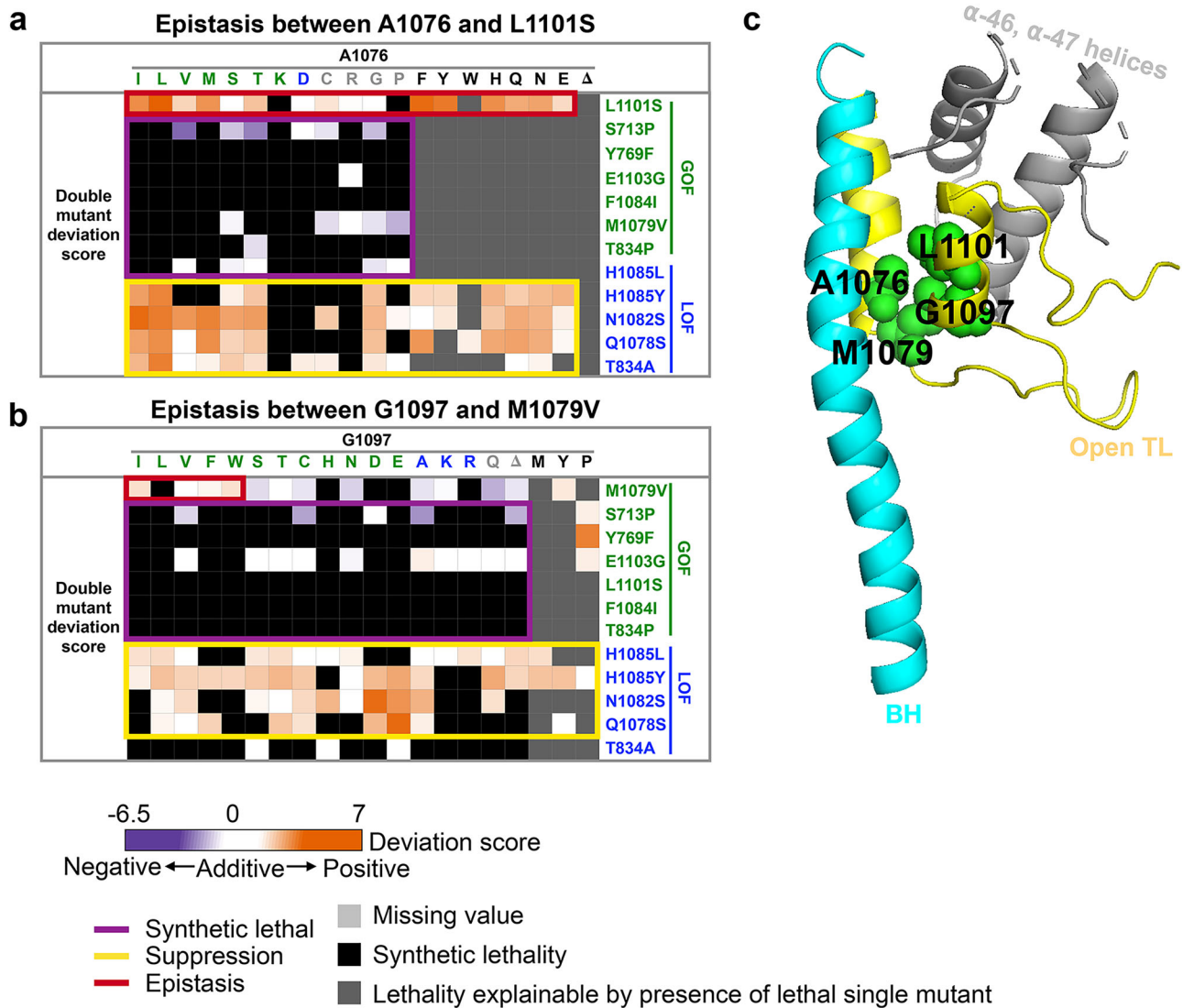
**Fig. 4 | Pol II TL interaction landscape distinguishes mutants with similar phenotypes.** **a** Design of the targeted double mutant libraries. All possible substitutions at each TL residue (represented with a simplified format in the left panel) and twelve “probe” mutations (eight within the TL and four in TL-proximal domains) (middle panel) were combined with to generate 7280 double mutants (right panel). 7276 mutants passed the reproducibility filter and were used for interaction analyses. **b** The percentage of functional interactions observed for each

probe mutant with viable GOF or LOF TL substitutions. Epistasis and sign epistasis are labeled with colored lines. **c** Pol II-TL functional interaction landscape with interactions represented by deviation scores. The upper panel shows interactions of GOF probe mutants in combination with viable GOF or LOF TL substitutions. The lower panel shows interactions of combinations with LOF probe mutants. **d** Principal component analysis (PCA) of deviation scores across double mutant interactions for 12 probe mutants (see “Methods”).



**Fig. 5 | Interaction networks of selected probe mutants.** The TL is shown in a circle with WT residues and positions labeled. All 20 substitutions of each TL residue are represented by a magenta arc under each WT residue, with tick marks representing individual substitutions at that position, and are colored by mutant

class. Comparison of interaction networks between S713P and T834P (a), E1103G and F1084I (b), and H1085Y and H1085L (c) showed the differences are significant ( $P < 0.0001$ ). The comparisons were performed with the Kruskal–Wallis test with  $P$  value correction with Dunn’s multiple comparisons test (Supplementary Table 7).



**Fig. 6 | Pol II TL interaction landscape reveals functional dependency of proximal residues.** Specific epistatic interactions observed between hydrophobic residues A1076 and L1101 (a), and M1079 and G1097 are shown as heatmaps (b). The x-axis of both heatmaps are 20 substitutions ordered by predicted phenotypic

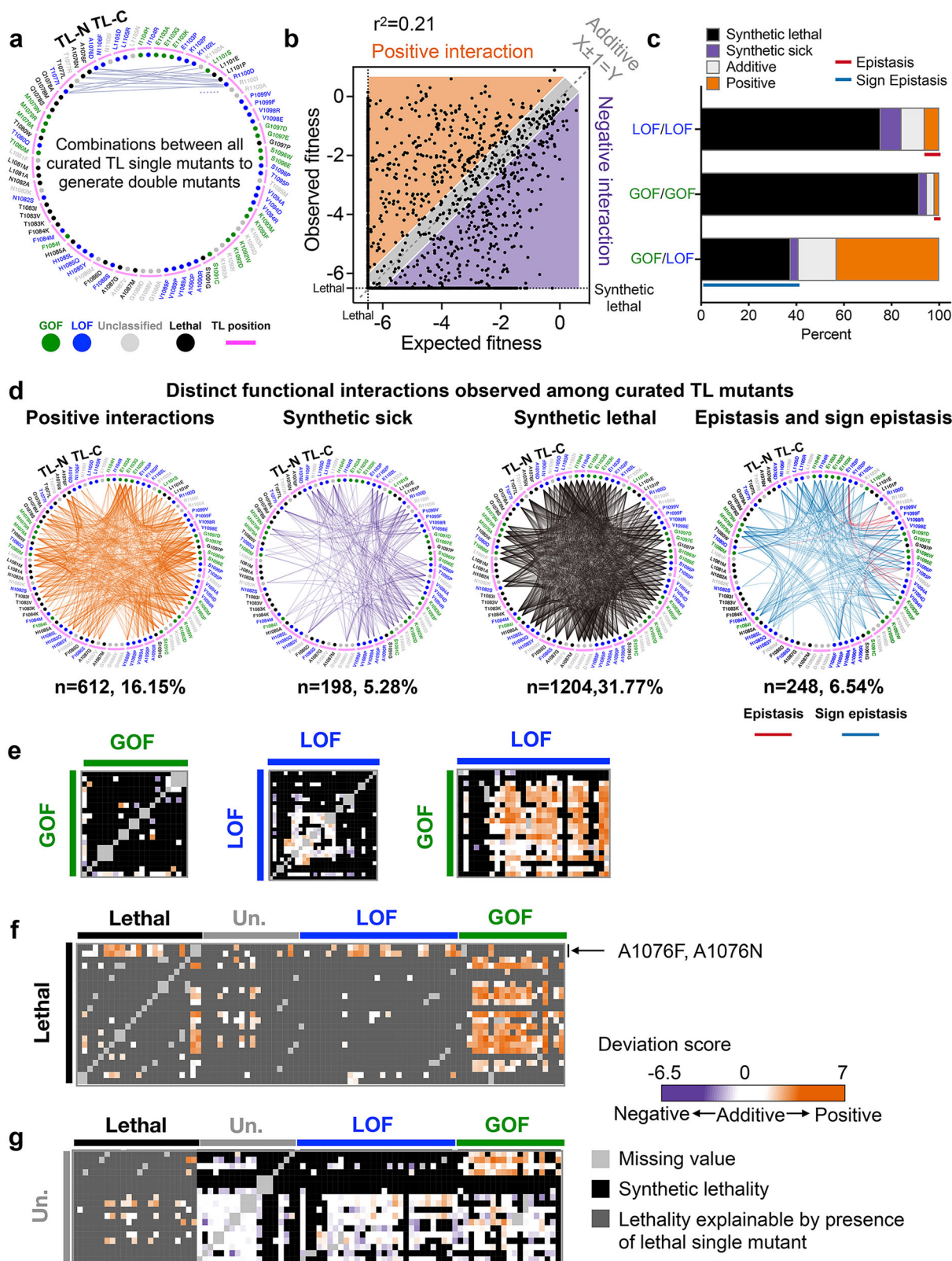
classes and color coded accordingly. The probe mutants are along the y-axis. c The epistatic interactions we identified between A1076 and L1101, as well as M1079 and G1097, are shown on the five-helix bundle of Pol II active site (PDB: 5C4X).

observed (Supplementary Fig. 9b). For example, lethal A1076 substitutions were suppressed by LOF probe mutants, suggesting their severe defects result from strong GOF status. However, some A1076 substitutions were also suppressed by the GOF probe mutant L1101S, indicating allele-specific mutual suppression or sign epistasis, consistent with A1076-L1101 interactions described above (Supplementary Fig. 9c). Similarly, ultra-sick/lethal LOF substitutions were commonly suppressed by GOF mutants, but some were suppressed by specific LOF mutants (Supplementary Fig. 9b). For example, S1091G was suppressed by most GOF mutants and specifically by the LOF V1094D (Supplementary Fig. 9c), suggesting allele-specific compensation. Furthermore, some lethal mutants exhibited interactions with specific GOF or LOF probe mutants. For example, F1084R was specifically suppressed by LOF probe Q1078S and GOF probe Y769F (Supplementary Fig. 9c). Notably, F1084, Q1078S, and Y769 are in close proximity in the substrate-bound, catalytic-favoring TL conformation. These types of allele-specific interactions potentially allow the TL and adjacent residues to evolve and differentiate while maintaining essential functions.

We note that strong epistasis is much more prevalent in the Pol II system than in other quantified proteins<sup>56,86–89</sup> (Supplementary Fig. 9d), likely due to the high rate of suppressive interactions stemming from Pol II mutants' opposing effects on catalysis.

### Mapping TL-internal residue interaction networks suggests widespread epistasis within the TL

To determine TL-internal interaction networks, we selected 2–4 substitutions for each TL residue, representing diverse phenotypes (GOF, LOF, lethal, or unclassified), and combined them with the selected substitutions across all other TL positions. This curated set of 3790 double mutants captures potential interactions between any two TL residue positions (Fig. 7a). Observed fitness of these double mutants was compared to predictions from the additive model, revealing widespread deviation from the model (i.e., large interactions) ( $r^2 = 0.21$ ), much smaller than the  $r^2$  (about 0.65–0.75) reported in other studies<sup>56,72,86,87,89</sup> (Fig. 7b), likely reflecting our selection criteria for mutations with diverse phenotypes. In general, mutations of the same class were additive or exacerbating, with some exceptions that



reveal functional dependencies between residues (Fig. 7c, e, Supplementary Fig. 10). Additionally, the strongest deviations from expectations were between mutations of opposing phenotypic classes (GOF + LOF combinations). Here, we commonly observed suppression as predicted for mutations acting independently in 43% of GOF + LOF combinations, but we also observed unexpected lethality or exacerbation (sign epistasis) for 41% of GOF + LOF combinations. Distinct

patterns of sign epistasis for GOF/LOF combinations provide the ability to distinguish functional requirements for GOF alleles to be able to confer GOF phenotypes (discussed below). Interactions were distributed throughout the TL, covering every residue, supporting connectivity across the TL (Fig. 7d). Notably, epistasis clustered within the C-terminal TL helix and adjacent regions (The far-right panel of Fig. 7d), supporting functional dependencies of TL-C terminal residues

**Fig. 7 | Widespread epistasis in the Pol II TL interaction landscape.** **a** Design of the pairwise double mutant library. We curated 2–4 substitutions for each TL residue (in total 90 substitutions,  $n(\text{GOF}) = 18$ ,  $n(\text{LOF}) = 30$ ,  $n(\text{Unclassified}) = 19$ ,  $n(\text{Lethal}) = 23$ ) and combined them with each other to generate 3910 double mutants. 3790 of them passed the reproducibility filter. WT TL residue positions are indicated with an arch. Phenotype classes of single substitutions are shown with corresponding colors. **b** An  $xy$ -plot of observed double mutant growth fitness measured in our experiment ( $y$ -axis) and expected fitness from the addition of two constituent single mutants' fitnesses ( $x$ -axis).  $N$  (positive) = 612.  $N$  (Negative) = 1402.  $N$  (Additive) = 1776.  $N$  (Sum) = 3790. Lethal threshold ( $-6.5$ ) is labeled with dotted lines on the  $x$ - and  $y$ -axis. The additive line where  $x \pm 1 = y$  is indicated by dashed line centered in the gray section. Simple linear regression was performed, and the best fit equation is  $y = 0.52x - 2.55$ ,  $r^2 = 0.21$ ,  $P < 0.0001$ . **c** Percent of interactions observed.  $N$  (LOF/LOF) = 412.  $N$  (GOF/GOF) = 156.  $N$  (GOF/LOF) = 534. **d** Various

groups of interactions are displayed in a network format for clarity. In these networks, the TL structure is simplified to a circle, with the TL-N terminal and C-terminal of the TL connected. WT TL residues are represented by arches, and selected substitutions are labeled beside these arches, with phenotypes indicated by colors. Lines connect two substitutions if there is interaction between them. For example, in the "Epistasis and sign epistasis" network, we selected three substitutions at K1102: K1102P, K1102L, and K1102A. They show epistasis with V1094D and R, T1095P, S1096P, V1098R, and P1099F and V. These interactions are labeled. The observed epistasis, accumulated at the TL-C terminal, indicates a functional dependency. **e–g**. The intra-TL functional interaction heatmaps of various combinations. Double mutant deviation scores are shown in the heatmap. GOF/GOF, LOF/LOF, and GOF/LOF combinations are shown in **(e)**. Combinations with lethal single substitutions are in **(f)**. Combinations with unclassified mutants are in **(g)**.

on one another, consistent with their proposed function in collaboratively stabilizing the open TL.

### Genetic interactions reveal insights into lethal or unclassified individual mutants

Most lethal mutants could be suppressed by at least one predicted GOF mutant (Fig. 7f, Supplementary Fig. 10), suggesting their lethality likely comes from reduced activity (LOF) below a viable threshold. However, this suppressibility also suggests that individual lethal substitution activity defects are mostly close to the viable threshold. This suggests that the contributions of most individual residues represent only a portion of total TL activity, which is expected to provide a  $\sim 1000$ -fold increase in catalytic activity<sup>90–92</sup>. Suppressibility of lethal residues by GOF alleles is also consistent with the greater probability of lethal substitutions being LOF rather than GOF. However, two lethal mutations at A1076 were suppressed by most LOF mutations or specific lethal mutants, but not GOF mutants, implying that their lethality resulted from being GOF. Additionally, unclassified single mutants generally showed limited interactions with GOF, LOF, or lethal classes, consistent with small or mostly independent effects. However, a few unclassified mutants showed suppression in combination with GOF mutants, suggesting they are either potentially atypical LOF not detected by phenotypic analysis or allele-specific interactions (Fig. 7g, Supplementary Fig. 10).

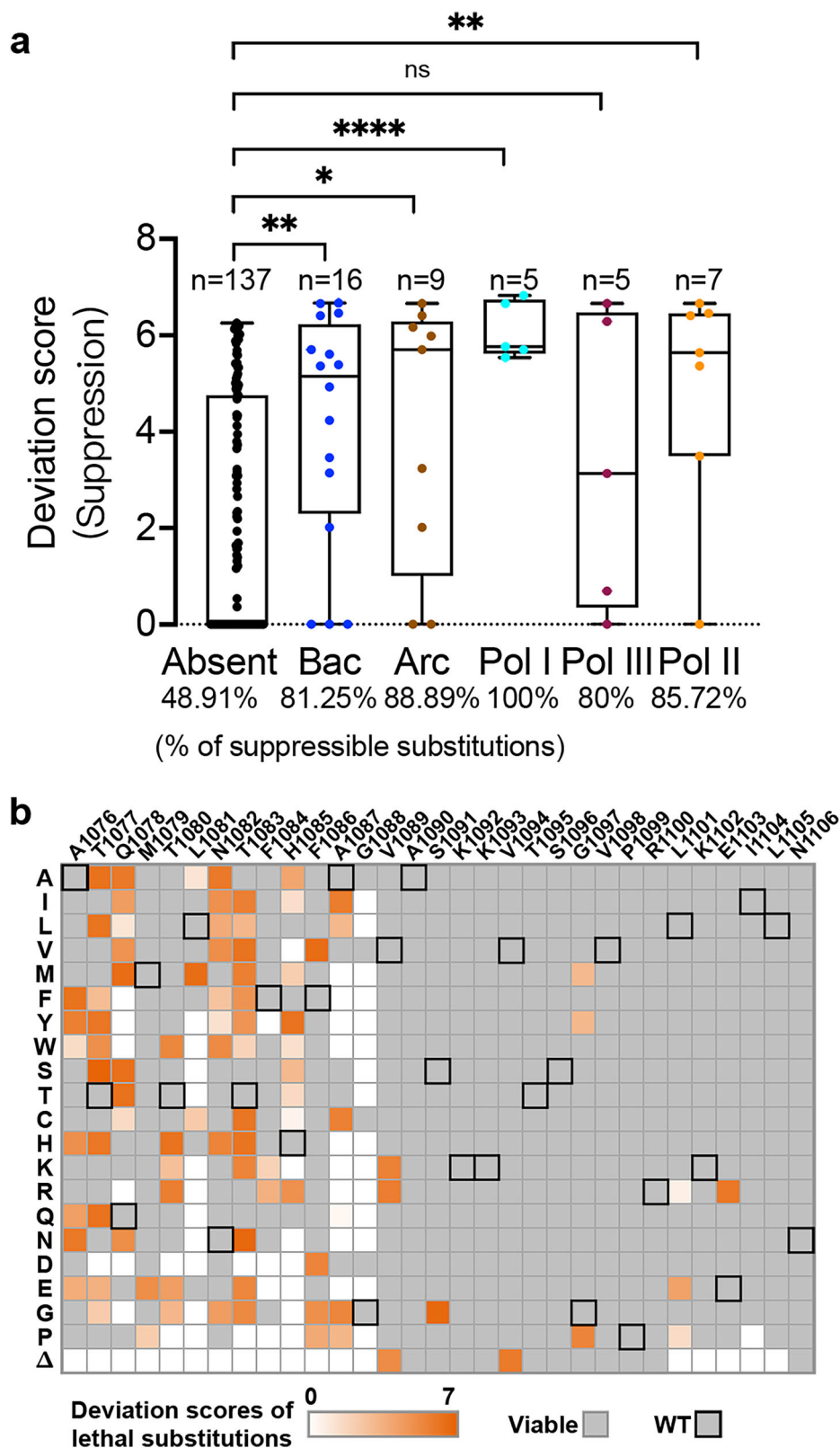
We reasoned that evolutionarily observed lethal substitutions would be closer to viability than those not observed and would therefore be more likely to be suppressible by Pol II GOF mutants. To determine this, we compared the suppressibility of substitutions observed in existing msRNAPs that are lethal but not present in extant msRNAPs by analyzing the maximum positive deviation scores from all double mutants containing individually lethal substitutions. Singly lethal substitutions (single substitutions that are lethal on their own in yeast Pol II) identified in our msRNAP TL MSA showed higher maximum deviation scores than substitutions lethal on their own in yeast and not present in the MSA (Fig. 8a, Supplementary Fig. 5b). Consistently, most L1081 substitutions cannot be suppressed, whereas a L1081M, (commonly a WT residue in bacteria and archaeal RNAPs) showed the highest suppressibility (Fig. 8b). These results indicate that singly lethal mutants in *S. cerevisiae* Pol II observed elsewhere msRNAP evolution, on average, maintain greater functionality and/or are more easily suppressed by single additional mutants in yeast Pol II. Interestingly, most lethal substitutions in A1087 and all lethal substitutions in G1088 cannot be suppressed (Fig. 8b), suggesting the strict restriction at these positions, consistent with the structural observation that A1087 and G1088 are in a small pocket between the bridge and funnel helices<sup>12,17</sup>. The TL has been estimated to enhance catalytic activity by 500–1000 fold<sup>90,91,93</sup>, while yeast viability tolerates only  $\sim 10$ -fold effects<sup>36</sup>. We conclude that lethal mutants observed as functional residues in other species are likely closer to the viability threshold, as

might be predicted to result from a series of small adaptive steps that allow their function.

### Coevolution of Pol II active site residues forms distinct sectors within specific domains and extends to the surfaces of Rpb1 and Rpb2

Our analyses suggest that even a highly conserved domain like the Pol II TL can be tightly integrated through coevolution of less conserved residues and that substitution tolerance is shaped by widespread epistasis. Changes in residue identities alter interaction networks, shaping the Pol II active site across evolution. To identify the coevolving residue networks in the Pol II active site, we employed statistical coupling analysis (SCA) for residues within Rpb1 and Rpb2, the two largest subunits of Pol II, which include the entire active site. SCA is well-suited for identifying coupled residue sectors, representing putative residue communication networks<sup>94–96</sup>. From 283 linked Rpb1/Rpb2 sequences, we identified 53 coevolving sectors (Supplementary Fig. 11). Residues within the active site form four major sectors (Fig. 9a–e). First, sector 37 includes many Rpb2 link domain residues and most TL residues, including those at the NIR and some at the TL C-terminal (Fig. 9b). This sector tightly surrounds the substrate binding pocket (Fig. 10a), befitting critical roles in catalysis. Second, sector 15 consists of four BH residues, most residues in switch 1 (part of the clamp domain), and many cleft residues, including the  $\alpha 46$ -47 loop, which accommodates incoming DNA<sup>8,13</sup> (Fig. 9c). Residues in this sector directly interact with Pol II subunits Rpb5 and Rpb6 (Fig. 10b), consistent with Rpb5's role in interacting with the incoming DNA<sup>97</sup>. While the  $\alpha 46$ -47 loop forms a five-helix bundle with two TL helices and the BH helix<sup>41</sup> (Fig. 6c), coupling analysis indicates it is primarily evolutionarily coupled with cleft residues instead of the TL. This suggests that the  $\alpha 46$ -47 loop may influence TL function by constraining the space for TL movement, rather than directly interacting through specific residue identities. Third, sector 1 involves active site residues surrounding the  $\text{Mg}^{2+}$  at the catalytic site and most F-loop residues (Fig. 9d). This sector forms a potential communication path that wraps around the TL and BH, connects to the BH through two BH residues, and extends to Rpb9, itself a subunit that forms important inputs to the TL and is required for transcription fidelity<sup>49,98,99</sup>. This pathway likely mediates communication between Rpb9 and the Pol II active site (Fig. 10c). Fourth, sector 2 contains most BH residues are coupled with the Rpb2 Fork and hybrid binding domain, forming a network linking the BH to Rpb2 residues (Fig. 9e). This network connects to Rpb5, which interacts with the downstream DNA, Rpb6, a shared subunit in three eukaryotic msRNAPs, Rpb9, and Rpb3 and Rpb10, which function in Pol II assembly<sup>100</sup> (Fig. 10d), representing a residue communication path potentially tied to various transcription activities.

Notably, coupling is not restricted to residues that are within or near the active site. Some sectors extend from the active site to the surface of Rpb1/Rpb2 (Supplementary Fig. 12a–f), suggesting that active site function, potentially through the TL, could be modulated by



allosteric interactions through the binding of transcription factors. Residues at interfaces with elongation factors can be coupled with those factors, yet some residues at interfaces can be coupled with residues distal to the interface, suggesting potential pathways for allosteric interactions. Examples of these are shown for TFIS (Supplementary Fig. 13), LEO1 (Leo1), ELOF1 (Elf1), and SUPT5H (Spt5) (Supplementary Fig. 14). Together, our findings highlight potential

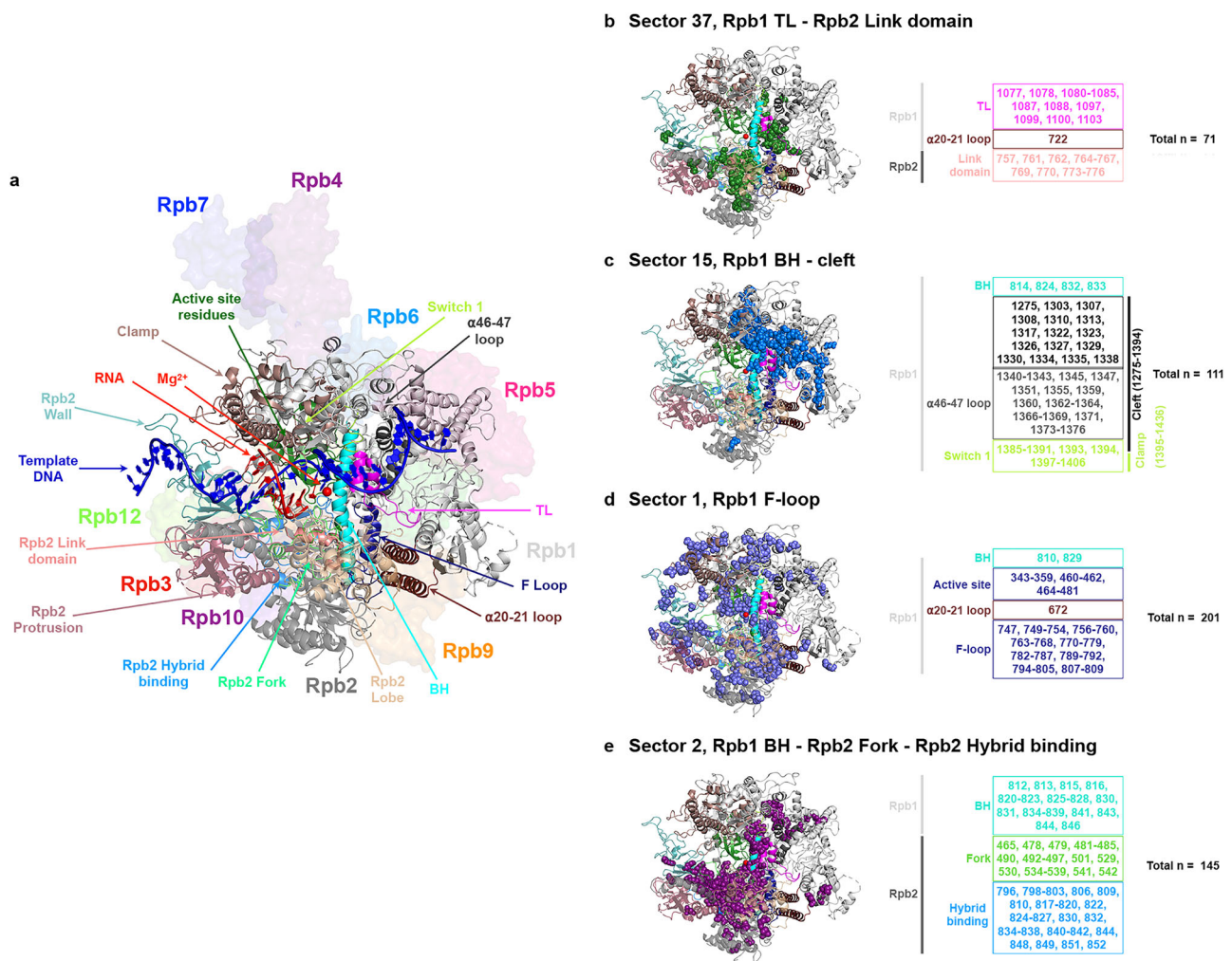
communication networks through evolutionary coupling that may underlie regulatory paths from Pol II surface to active site.

## Discussion

How individual mutants alter a protein's function is not necessarily straightforward at the mechanistic level. Amino acid substitutions both remove functionality of the WT residue but replace that

**Fig. 8 | Contextual epistasis shapes TL evolution.** **a** A comparison of the maximum deviation score of each TL lethal single substitution that is present in any evolutionary TL haplotypes from bacterial, archaeal, or eukaryotic Pils versus those that have not been observed in any species. The evolutionary TL haplotypes were from multiple sequence alignments (MSAs). Nine substitutions were found in an MSA of 542 archaeal TL sequences that are lethal when present in yeast as a single substitution. Seventeen were found in an MSA of 1403 bacterial TLs, five were found in 749 Pol I TLs, seven were found in 499 Pol II TLs, and five were found in 539 Pol III TLs. Evolutionarily observed lethal substitutions were compared to those unobserved in our TL MSA. The percentage of total suppressible lethal single mutants for each group is labeled at the bottom of the plot. Boxes are: center line,

median; box limits, second and third quartiles; whiskers, maximum and minimum points. Statistical comparison was done with the two-sided Mann–Whitney test, and the significant levels (Bacteria vs. Absent,  $P = 0.0011$ . Archaea vs. Absent,  $P = 0.0122$ . Pol I vs. Absent,  $P < 0.0001$ . Pol III vs. Absent,  $P = 0.1277$ . Pol II vs. Absent,  $P = 0.0018$ ) are shown in the figure. **b** A heatmap displays the highest deviation scores of lethal single substitutions. The  $x$ -axis represents the wild-type residues and positions of the yeast Pol II TL, while the  $y$ -axis shows 20 substitutions. Viable substitutions are marked with gray boxes, and wild-type amino acids are indicated with gray boxes with a black edge. The highest deviation scores of lethal substitutions are shown in colored boxes, ranging from white to orange, reflecting the strength of suppression.



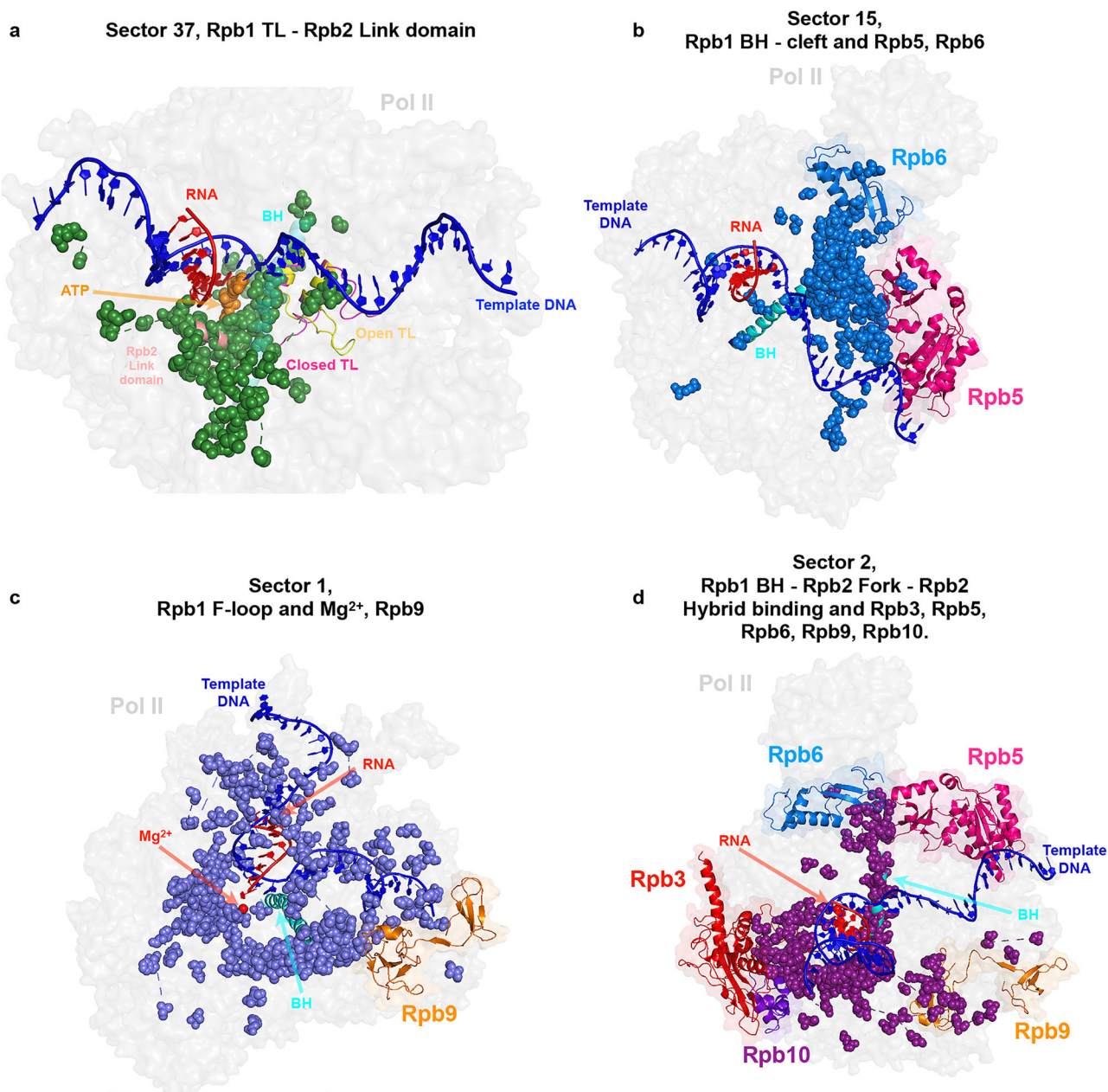
**Fig. 9 | Ultra-conserved TL co-evolves with Pol II residues through diverse pathways.** **a** Structure representations of Pol II active site. Pol II subunits are shown with the surface in different colors. The active site domains are in cartoon format. Rpb1 active site residues 346–375 and 436–508. Rpb1 TL, 1076–1106. Rpb1 BH, 810–846. Rpb1 α20-21 loop, 672–737. Rpb1 F loop, 747–810. Rpb1 α46-47 loop, 1339–1378. Rpb1 switch 1, 1385–1407. Rpb1 clasp, 1–95, 235–346, and 1395–1436.

Rpb2 link domain, 757–776. Rpb2 Fork, 465–547. Rpb2 hybrid binding, 750–852. Rpb2 wall, 852–973. Rpb2 protrusion, 44–218. Rpb2 lobe, 218–405. **b–e** Four major coevolution sectors of the Pol II active site residues were identified through statistical coupling analysis. The residues within each active site domain and the total number of residues in each sector are labeled. These sectors are mapped on the yeast Pol II Rpb1/Rpb2 structure (PDB: 5C4X).

functionality with something different. By altering the local environment within a protein or potentially propagating effects to distant locations, each substitution potentially can be quite different. These differences may not be apparent as phenotypic outputs, and phenotypic assays may not have granularity to distinguish different biophysical behaviors if they result in similar outputs. For Pol II mutants, even high-resolution phenotypic analyses, such as gene expression profiling or genetic interaction profiling between Pol II mutants and deletions in other yeast genes<sup>52</sup>, suggest that LOF and GOF mutants represent a

continuum of defects that match enzymatic activity in vitro. Therefore, these profiles also appear dependent on the output of Pol II activity defects and cannot distinguish potential differences in the underlying mechanism.

Through systematic detection of genetic interactions within the Pol II active site, we have identified functional relationships between amino acids across the TL and between TL substitutions and others. In the absence of double mutant epistasis analyses, it would not be possible to differentiate similar alleles from one another. L1101S and



**Fig. 10 | Coevolutionary sectors of Pol II active site residues extend to the surface of Pol II, forming connections with other Pol II subunits. a** The coevolutionary sector of Rpb1 TL and Rpb2 link domain residues surrounds the substrate ATP, indicating its potential role in catalysis. The sector is mapped on the yeast Pol

II structure with TL in open state (PDB: 5C4X) and closed state (PDB: 8U9R). **b** Residues in sector 15 connect with Rpb5 and Rpb6. **c** Residues on sector 1 connect with  $Mg^{2+}$  and Rpb9. **d** Residues in sector 2 connect with Rpb5, Rpb6, Rpb3, Rpb9, and Rpb10. These sectors are mapped on the yeast Pol II structure (PDB: 5C4X).

E1103G, for example, are two GOF alleles very close to each other in Pol II structure and confer similar phenotypic landscapes across various growth conditions. Here, we find that their distinct interactions support that substitutions at I101 and I103 target distinct residue networks (Figs. 4c and 6a, Supplementary Figs. 6 and 7). L1101 functions in the TL hydrophobic pocket while E1103 interacts and co-evolves with a number of TL-external residues that together support interactions that maintain the open TL conformation. We also observed connections between TL C-terminal residues that suggest a limit to how disruptions to structure there can alter Pol II activity (Fig. 7d). Helix-disrupting LOF proline substitutions in at least two TL positions showed epistasis with multiple substitutions in the back of the TL (I094–I098), suggesting that their functions require TL C-terminal helix structure and in the

absence of that structure (proline disruption) effects are no longer additive.

The strongest epistatic interactions were observed between two pairs of hydrophobic residues, A1076 and L1101, and M1079 and G1097 (Fig. 6), suggesting their interdependence in maintaining TL open conformation stability<sup>41</sup>. Considering the TL's multiple active site roles, the observed interaction networks may also contribute to other NAC steps or off-pathway steps (e.g., pausing or backtracking). Structural studies have proposed that these hydrophobic residues form a pocket stabilizing the TL by bundling two TL helices with the BH and  $\alpha$ 46-47 loop in a five-helix bundle<sup>28,41</sup>. However, our coupling analysis suggests that the  $\alpha$ 46-47 loop is evolutionarily coupled with cleft residues rather than the TL, indicating the  $\alpha$ 46-47 loop may indirectly impact TL

function by limiting the space available for TL movement, rather than through direct interactions with TL residues (Fig. 9c).

Elongation factors bind Pol II and alter its activity, but the mechanisms by which they do so are not known<sup>50,101</sup>. We observed a high level of genetic interactions between residues outside the TL and residues within it, including allele-specific reshaping of TL mutant space upon single substitution outside the TL (Fig. 4). The fact that minor mutational changes outside the TL can apparently functionally perturb the TL would be consistent with the idea that minor alterations to Pol II structure upon elongation factor binding could easily propagate into the active site via the TL or the BH. As an example, human Rtf1 has been observed to project a domain into the Pol II structure adjacent to the BH (in yeast, this region is occupied instead by Rpb2<sup>102</sup>). These contacts have been proposed to alter Pol II activity. We would propose that the paths for such alteration to activity may follow the coupling sectors we have observed by SCA (Fig. 9).

How different individual substitutions are under the surface is critical for understanding plasticity in protein mechanisms and how they might be altered by evolutionary change. A key open question in nucleic acid polymerase mechanisms is the paths for protons in the reaction (for example, deprotonation of the synthesized strand 3'-OH and protonation of pyrophosphate leaving group, for example<sup>80,81,83,85,103-105</sup>). For msRNAPs, the association with the incoming NTP by a universally conserved histidine led to the proposal that this residue might donate a proton during the reaction<sup>12,82,103</sup>. Some substitutions at this position can provide minimal essential function (e.g., tyrosine, arginine), while others are only moderately defective (glutamine). Unexpectedly, we found that H1085L was very well tolerated for growth<sup>17</sup> and the Landick lab has proposed that this substitution supports catalysis through positional but not chemical effects<sup>84,85</sup>. Our studies here were quite surprising in that they indicated that L1085 Pol II has unique behavior when perturbed by all possible TL substitutions and is entirely distinct from H1085Y (where we have direct observations of all possible intra-TL doubles) or H1085A or H1085Q (curated doubles) (Fig. 5c, Supplementary Fig. 15). These residue specific behaviors suggest that each substitution may have different properties, and compatibility with function may not necessarily represent similar function under the surface.

Evolutionary change over time can alter protein function, but it can also alter protein functional plasticity. Recent work from the Thornton lab elegantly demonstrates that phenotypes of substitutions to residues conserved over hundreds of millions of years can change over evolutionary time and can do so unpredictably and transiently during evolution<sup>61</sup>. msRNAPs have structures and functions conserved over billions of years, and deep within their active sites is a mobile domain, the TL, that has large functional constraints on its sequence. The TL sequence must be able to fold into multiple states and maintain recognition of the same substrates across evolutionary space, and it shows high identity even between distantly related species. Here, we show that the TL, and likely the entire Pol II active site, exhibits a great amount of plasticity through non-conserved positions that are essential for compatibility of the TL and surrounding domains. A limitation of our study is its reliance primarily on genetic data and SCA. Future biochemical and biophysical studies will give mechanistic insight into the activity defects underlying the functional connections we observe, for example, connecting kinetic properties or structural states to specific mutants. Nonetheless, our results illustrating widespread epistasis and allele-specific effects of single and double mutants predict that comparative analyses among Pol I, II, and III will reveal shared but also enzyme-specific mechanisms due to higher-order epistasis shaping the functions of conserved residues.

## Methods

Strains and plasmids presented in this article will be provided upon request.

## Design and Synthesis of TL mutant libraries

We updated and extended the fitness dataset of Qiu et al.<sup>17</sup> using a similar methodology, but with adjusted conditions and a second-generation mutant library strategy to generate a complete Pol II TL mutation-phenotype map and examine genetic interactions. Mutants were constructed by oligo synthesis (Agilent) and screened for phenotypes previously established as informative for Pol II mutant biochemical defects. Programmed oligonucleotide library pools included all 620 single TL residue substitutions and deletions for Rpb1 amino acids 1076–1106 (Library 1), 3914 pairwise double substitutions (Library 2), 4800 targeted double substitutions (Library 6), and 3373 multiple substitutions (Library 3–5), along with the WT *S. cerevisiae* Pol II TL allele at a level of ~15% of the total variants, enabling precise quantification (see Supplementary Tables 1 and 3). Each synthesized region contained a mutated or WT Pol II TL sequence and two flanking regions at the 5' and 3' ends of the TL-encoding sequence. These flanking regions also contained designed “PCR handle” (20 bp) sequences, allowing distinct subsets of oligos to be amplified from synthesized pools using selected primers for PCR, and additional flanking WT Pol II sequences allow for further extension of homology arms by PCR “sewing” (details are in Supplementary Method 3).

## Introduction of libraries into yeast

Synthesized mutant pools were transformed into yeast (CKY283) along with an *RPB1*-encoding plasmid where the TL-encoding sequence was replaced with an *MluI* restriction site for linearization as described in Qiu et al.<sup>17</sup>. This strategy allows construction of *rpb1* mutant libraries by gap repair between library fragments and the linearized vector. Briefly, the synthesized oligo pools were amplified by limited cycles of emulsion PCR to limit template switching. Extension of flanking homology arms of ~200 bp was added by PCR sewing. Amplified TL sequences with extended flanking regions were co-transformed with linearized pRS315-derived *CEN LEU2* plasmid (pCK892) into CKY283, allowing gap repair via homologous flanking regions. In our experience, phenotypes of integrated alleles and single copy plasmids are qualitatively very similar, but this has not been systematically compared quantitatively. To detect potential residue-residue interactions between the TL and TL-proximal domains including the Rpb1 BH, Funnel Helix alpha-21 and Rpb2, the Pol II TL single mutant pool (Library 1, 620 mutant alleles and 111 WT alleles) was co-transformed individually with gapped plasmids encoding an additional *rpb1* allele T834P (Library 7), T834A (Library 8), or Funnel Helix alpha-21 S713P (Library 10) into CKY283 respectively, or with the gapped WT *RPB1* plasmid into a strain with the genomic mutation, *rpb2* Y769F (Library 9). These co-transformations created double mutants between the TL and TL-proximal mutants. The WT allele in the single mutant pool represented the single probe mutant due to substitutions outside the TL on the plasmid or in the strain background. To distinguish between a fully WT TL and a WT TL representing the TL of a mutant allele elsewhere, a WT Pol II TL allele with a silent mutant at T1083 (WT codon ACC was replaced with ACT) was co-transformed with a plasmid containing gapped WT *RPB1* in a WT strain in parallel. 15% of the transformants with silent mutation were mixed with transformants of double mutants. The silent mutation allowed us to distinguish the WT and the single mutants.

## Phenotyping libraries with different conditions

Each transformation was done in three biological replicates. After transformation, Leu<sup>+</sup> colonies were collected from **SC-Leu** plates and replated on **SC-Leu+5FOA** to select for cells having lost the *RPB1 URA3* plasmid and to measure growth defects. 5FOA-resistant colonies were then scraped and replated on different conditions to assess functional differences between WT, GOF, and LOF mutants<sup>17,37</sup>. **SC-Leu+20 µg/L MPA** (Fisher Scientific 150676121) evaluates transcription start site (TSS) shifting. MPA depletes GTP levels in yeast. To generate GTP, *IMD2* expression is required, but its promoter has multiple

transcription start sites (TSS)<sup>106,107</sup>. Transcripts from the downstream TSS express functional *IMD2* upon MPA treatment. WT and LOF mutants are resistant to MPA due to their ability to use downstream TSSs for *IMD2* expression, while GOF mutants are sensitive as they can only use upstream TSSs and fail to express functional *IMD2*<sup>52,76</sup>. **SC-Leu+15 mM Mn** (Sigma): Mn<sup>2+</sup> compromises transcription fidelity<sup>108</sup> and likely also increases Pol II catalysis; while WT and LOF mutants are resistant, GOF mutants are sensitive either due to decreased fidelity or exacerbated catalytic increase. **YPRaf** and **YPRafGal**: the *gal10Δ56* allele compromised *GAL10* 3'-end formation and polyadenylation, leading to transcription read-through that inhibits the downstream *GAL7* gene<sup>109,110</sup>. The transcription read-through in WT *gal10Δ56* cells leads to the accumulation of a toxic galactose metabolite because of defects in *GAL7* expression. Some GOF and most LOF mutants show varying galactose resistance, likely through increasing termination in the *GAL10-GAL7* intergenic region or otherwise reducing read-through interference with *GAL7*. YPRaf serves as a control for YPRafGal to demonstrate that the defect in YPRafGal is galactose toxicity. **SC-Lys** tests lysine auxotrophy. The *lys2-128Δ* allele in the strain background has a *Ty1* transposable element insertion, which prevents normal expression of *LYS2*<sup>111</sup>. WT and LOF mutants are Lys<sup>-</sup>, while GOF mutants utilize a normally silent promoter within the *Ty1*  $\delta$  element insertion, becoming Lys<sup>+</sup>. **SC-Leu+3% formamide** (Fisher Scientific AM9342), formamide destabilizes protein hydrogen bonds<sup>112</sup>, with most mutants showing resistance except for slight sensitivity in LOF mutants. Details of cell numbers plated on each plate and screening time of each plate are in Supplementary Table 5. Details of the high-efficiency transformation protocol are in Supplementary Method 1.

### Generation of libraries for quantification by amplicon sequencing

Genomic DNA of each screened library was extracted using the Yeastar genomic DNA kit according to the manufacturer's instructions (Zymo Research D2002). To ensure adequate DNA for sequencing, the TL regions of all libraries were amplified with PCR cycles that were verified to be in the linear range by qPCR to minimize disturbance of allele distributions, and under emulsion PCR conditions (EURx Micellula DNA Emulsion & Purification (ePCR) PCR kit 3600-02) to limit template switching. Details are in Supplementary Methods 2 and 3. To multiplex samples, we employed a dual indexing strategy wherein 10 initial barcodes for differentiating 10 mutant libraries were added during the initial amplification using 10 pairs of custom primers. In a second amplification, 28 primers containing 28 NEB indices were used to add a second index for distinguishing conditions and replicates (NEBNext Multiplex Oligos for Illumina, NEB E6609S) (see Supplementary Table 4). As a result, a sample-specific barcode was present for each set of variants. The indexed, pooled samples were sequenced by single-end sequencing on an Illumina Next-Seq (150nt reads). On average, over 11 million reads were obtained for individual samples with high reproducibility from two rounds of sequencing.

### Data cleaning and fitness calculation, and normalization

Reads of mutants were sorted into appropriate libraries and conditions by detecting particular indices after sequencing. Read counts were estimated by a codon-based alignment algorithm to distinguish reads that exactly matched the designated codons of mutants<sup>113</sup>. To clean the data, mutant reads with coefficients of variation greater than 0.5 in the control condition (SC-Leu) were excluded from the analysis. The mutant read count was increased by 1 to calculate the allele frequency under different conditions. To measure and compare the phenotypes of all mutants, the mutant phenotypic score (fitness) was calculated by allele frequency change of a mutant under selective conditions relative to the unselective condition compared to the frequency change of WT. The formula for

calculating fitness is shown below.

$$\text{Fitness}(\text{mut}) = \log \left[ \frac{f^{\text{mut, sele}}}{f^{\text{mut, unsele}}} \right] - \log \left[ \frac{f^{\text{WT, sele}}}{f^{\text{WT, unsele}}} \right] \quad (1)$$

We applied min-max normalization to bring the median growth fitness of mutants measured for the ten variant libraries to the same level for direct comparison (formula is shown below). In each library, we divided mutants into several groups based on allele counts in the control condition. Mutants with read count differences of less than 10 are present in one group. The WT growth fitness was set as the maximum value, and the minimum fitness in each group was the minimum. Min-max normalization was used to equalize the growth fitness into the same range between various groups inside each library. Additionally, we utilized min-max normalization to level the mutant fitness across all ten libraries, with WT fitness as Max and minimal fitness in each library as the minimum. As a result, mutant growth fitness was scaled to one range and could be used to determine genetic interactions.

$$X' = \frac{X - X \text{ min}}{X \text{ max} - X \text{ min}} \quad (2)$$

### Determination of functional interactions

The genetic interactions between single substitutions were determined by comparing the multiple-substitution mutant normalized median fitness to the log additive of the single-substitution normalized median fitness. The simplified formula is as follows:

$$\text{Deviation score}(M_1M_2M_3) = \text{Fitness}(M_1M_2M_3) - [\text{Fitness}(M_1) + \text{Fitness}(M_2) + \text{Fitness}(M_3)] \quad (3)$$

- $-1 < \text{Deviation score} < 1$ , the interaction among the constituent single mutants is additive, and mutants are acting independently.
- $\text{Deviation score} \geq 1$ , the interaction is non-additive and is positive, including suppression and epistatic interactions.
- $\text{Deviation score} \leq -1$ , the interaction is non-additive and is negative, including synthetic sick, synthetic lethal, and sign epistasis interactions.

Any mutation with fitness smaller than the lethal threshold ( $-6.50$ ) was classified as an ultra-sick/lethal mutant, and its fitness was normalized to  $-6.50$  for calculation of deviation scores. Synthetic sickness and synthetic lethality were distinguished by if a double mutant was viable or lethal (fitness is greater than or equal to the lethal threshold  $-6.5$ ) when two constituent mutations are viable. Synthetic lethality can be further classified into two types. First, additive synthetic lethality was determined when the expected double mutant fitness calculated by the additive model was lethal (expected fitness =  $-6.5$ ) and the observed double mutant fitness was also lethal (fitness =  $-6.5$ ) (in this case, the deviation score = 0). Second, the beyond-additive synthetic lethality was determined when the expected double mutant was viable (expected fitness  $> -6.5$ ) while the observed double mutant fitness was lethal (fitness =  $-6.5$ ) (in this case, the deviation score  $< 0$ ). To separate these two situations in our figures, we labeled additive synthetic lethality as black and beyond-additive synthetic lethality as purple. Details of formulas are in Supplementary Method 4.

### Mutant classification using two multiple logistic regression models

We trained two multiple logistic regression models to distinguish GOF and LOF mutants using phenotypic fitness from SC-Leu+MPA, SC-Lys, and YPRafGal conditions of 65 single mutants. This dataset included 20

mutants with transcription rates measured by biochemical experiments (13 GOF and seven LOF)<sup>17,37,52</sup>, and 45 additional mutants that were hierarchically clustered with verified mutants (12 GOF, 26 LOF, 1 WT, and 6 that were not GOF or LOF mutants)<sup>17</sup>. Both models incorporated intercept, main effects, and two-way interactions were involved in defining both models, with a cutoff threshold of 0.75 applied to classify mutants in both of the GOF and LOF models.

Model for predicting the probability of a mutant being a GOF:

$$y = \frac{1}{1 + e^{(1.816 + 2.542*fMPA - 1.942*fLys + 0.06566*fGal - 0.5297*fMPA*fLys - 0.08373*fMPA*fGal + 0.02556*fLys*fGal)}} \quad (4)$$

Model for predicting the probability of a mutant being LOF:

$$y = \frac{1}{1 + e^{(1.916 - 1.392*fMPA - 1.328*fLys - 0.8353*fGal - 0.01112*fMPA*fLys - 0.2992*fMPA*fGal + 0.8823*fLys*fGal)}} \quad (5)$$

Both models showed accuracy, with the area under ROC close to one (Supplementary Fig. 3a). The details are provided in Supplementary Table 6.

### Principal component analysis (PCA)

Deviation scores of curated and probe double mutants were analyzed in PCA. The scripts used R language v4.0.3 (<https://www.R-project.org/>) with R packages tidyverse v1.3.1 (<https://www.tidyverse.org>), prompt (stats v3.6.2 (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp>)), ggplot2 v3.3.3 (<https://ggplot2.tidyverse.org>), dplyr v1.0.6 (<https://dplyr.tidyverse.org>), and missMDA v1.18 (<https://dplyr.tidyverse.org>).

### t-SNE projection

Allele frequencies for all mutants in nine conditions with three replicates were analyzed by t-SNE (Perplexity = 50) or k-means (clusters = 20). Thirteen clusters with ultra-sick to lethal mutants as the majority were eliminated. The remaining mutants were analyzed again with t-SNE (Perplexity = 100) and k-means (cluster = 10). The scripts utilized R language v4.0.3 (<https://www.R-project.org/>), along with R packages Rtsne v0.15 (<https://github.com/jkrijthe/Rtsne>), ggplot2 v3.3.3 (<https://ggplot2.tidyverse.org>), and k-means (stats v3.6.2 (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans>)).

### Statistical coupling analysis

5787 eukaryotic homologous sequences of yeast Rpb1 from a published multiple sequence alignment (MSA) and 4000 Rpb2 sequences from the NCBI nr database were cleaned and used in the SCA. Sequence identity was reduced to 90% with the T-coffee package v12.00.7fb08c2<sup>114</sup>. Pol I, II, and III sequences were separated based on a phylogenetic tree constructed with FastTree 2<sup>115</sup>. Finally, 283 matched Pol II Rpb1 and Rpb2 sequences, with each matched pair sharing the same species name, were connected with 10 Ns and aligned with T-coffee. This newly generated MSA was used for SCA with the Python-based package pySCA v6.1<sup>95</sup>. The scripts were adapted from <https://github.com/ranganathanlab/pySCA>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw sequencing data have been deposited on the NCBI SRA (Sequence Read Archive) database under BioProject [PRJNA948661](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA948661). Processed mutant count, fitness, and processing codes are available on GitHub [[https://github.com/Kaplan-Lab-Pitt/TLs\\_Screening.git](https://github.com/Kaplan-Lab-Pitt/TLs_Screening.git)] and Zenodo [<https://doi.org/10.5281/zenodo.16370006>]. PDB structures 5C4X, 2E2H, and 8U9R were used to generate figures. Source data are provided with this paper.

### Code availability

The codes for calculating deviation scores, PCA, t-SNE projection, statistical coupling analysis, and generating figures are available in GitHub [[https://github.com/Kaplan-Lab-Pitt/TLs\\_Screening.git](https://github.com/Kaplan-Lab-Pitt/TLs_Screening.git)] and in Zenodo [<https://doi.org/10.5281/zenodo.16370006>]<sup>116</sup>.

### References

- Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **12**, 89–97 (2002).
- Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).
- Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599–610 (1985).
- Zhang, G. et al. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
- Vassilyev, D. G. et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).
- Hirata, A., Klein, B. J. & Murakami, K. S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**, 851–854 (2008).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).
- Fernandez-Tornero, C. et al. Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**, 644–649 (2013).
- Hoffmann, N. A. et al. Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**, 231–236 (2015).
- Malinen, A. M. et al. Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res.* **40**, 7442–7451 (2012).
- Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
- Kaplan, C. D. Basic mechanisms of RNA polymerase II activity and alteration of gene expression in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta Gene Regul. Mech.* **1829**, 39–54 (2013).

14. Dangkulwanich, M. et al. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *eLife* **2**, e00971 (2013).
15. Bar-Nahum, G. et al. A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
16. Weinzierl, R. O. The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain. *BMC Biol.* **8**, 134 (2010).
17. Qiu, C. et al. High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop. *PLoS Genet.* **12**, e1006321 (2016).
18. Da, L. T. et al. Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat. Commun.* **7**, 11244 (2016).
19. Mazumder, A., Lin, M., Kapanidis, A. N. & Ebright, R. H. Closing and opening of the RNA polymerase trigger loop. *Proc. Natl. Acad. Sci. USA* **117**, 15642–15649 (2020).
20. Wang, B., Predeus, A. V., Burton, Z. F. & Feig, M. Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophys. J.* **105**, 767–775 (2013).
21. Larson, M. H. et al. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **109**, 6555–6560 (2012).
22. Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P. & Thomm, M. The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.* **41**, 7048–7059 (2013).
23. Kireeva, M. L. et al. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol. Cell* **30**, 557–566 (2008).
24. Liu, X., Bushnell, D. A. & Kornberg, R. D. RNA polymerase II transcription: structure and mechanism. *Biochim. Biophys. Acta* **1829**, 2–8 (2013).
25. Kaplan, C. D. & Kornberg, R. D. A bridge to transcription by RNA polymerase. *J. Biol.* **7**, 39 (2008).
26. Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J. Biol.* **7**, 40 (2008).
27. Silva, D. A. et al. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. USA* **111**, 7665–7670 (2014).
28. Lin, G. et al. Structural basis of transcription: RNA polymerase II substrate binding and metal coordination using a free-electron laser. *Proc. Natl. Acad. Sci. USA* **121**, e2318527121 (2024).
29. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28**, 2663–2676 (2014).
30. Xu, L. et al. Dissecting the chemical interactions and substrate structural signatures governing RNA polymerase II trigger loop closure by synthetic nucleic acid analogues. *Nucleic Acids Res.* **42**, 5863–5870 (2014).
31. Vassilyev, D. G. et al. Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163–168 (2007).
32. Da, L. T., Wang, D. & Huang, X. Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J. Am. Chem. Soc.* **134**, 2399–2406 (2012).
33. Liu, B., Zuo, Y. & Steitz, T. A. Structures of *E. coli* sigmaS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc. Natl. Acad. Sci. USA* **113**, 4051–4056 (2016).
34. Seibold, S. A. et al. Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *Biochim. Biophys. Acta* **1799**, 575–587 (2010).
35. Malagon, F. et al. Mutations in the *Saccharomyces cerevisiae* RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics* **172**, 2201–2209 (2006).
36. Kaplan, C. D., Larsson, K. M. & Kornberg, R. D. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell* **30**, 547–556 (2008).
37. Kaplan, C. D., Jin, H., Zhang, I. L. & Belyanin, A. Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. *PLoS Genet.* **8**, e1002627 (2012).
38. Kireeva, M. L. et al. Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase. *BMC Biophys.* **5**, 11 (2012).
39. Windgassen, T. A. et al. Trigger-helix folding pathway and Sl3 mediate catalysis and hairpin-stabilized pausing by *Escherichia coli* RNA polymerase. *Nucleic Acids Res.* **42**, 12707–12721 (2014).
40. Nayak, D., Voss, M., Windgassen, T., Mooney, R. A. & Landick, R. Cys-pair reporters detect a constrained trigger loop in a paused RNA polymerase. *Mol. Cell* **50**, 882–893 (2013).
41. Barnes, C. O. et al. Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol. Cell* **59**, 258–269 (2015).
42. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
43. Sydow, J. F. et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).
44. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell* **57**, 408–421 (2015).
45. Kuldell, J. C. & Kaplan, C. D. RNA Polymerase II activity control of gene expression and involvement in disease. *J Mol Biol.* **437**, 168770 (2025).
46. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIIS. *Mol. Cell* **16**, 955–965 (2004).
47. Lennon, C. W. et al. Direct interactions between the coiled-coil tip of DksA and the trigger loop of RNA polymerase mediate transcriptional regulation. *Genes Dev.* **26**, 2634–2646 (2012).
48. Hein, P. P. et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
49. Kaster, B. C., Knippa, K. C., Kaplan, C. D. & Peterson, D. O. RNA polymerase II trigger loop mobility: indirect effects of Rpb9. *J. Biol. Chem.* **291**, 14883–14895 (2016).
50. Schier, A. C. & Taatjes, D. J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **34**, 465–488 (2020).
51. Leng, X. Y. et al. Organismal benefits of transcription speed control at gene boundaries. *EMBO Rep.* **21**, e49315 (2020).
52. Braberg, H. et al. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**, 775–788 (2013).
53. Tesileanu, T., Colwell, L. J. & Leibler, S. Protein sectors: statistical coupling analysis versus conservation. *PLoS Comput. Biol.* **11**, e1004091 (2015).
54. Phillips, P. C. The language of gene interaction. *Genetics* **149**, 1167–1171 (1998).
55. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
56. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016).
57. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).

58. Karageorgi, M. et al. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **574**, 409–412 (2019).
59. Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
60. Faure, A. J. et al. Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022).
61. Park, Y., Metzger, B. P. H. & Thornton, J. W. Epistatic drift causes gradual decay of predictability in protein evolution. *Science* **376**, 823–830 (2022).
62. Ding, D. et al. Co-evolution of interacting proteins through non-contacting and non-specific mutations. *Nat. Ecol. Evol.* **6**, 590–603 (2022).
63. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. USA* **99**, 14878–14883 (2002).
64. Lunzer, M., Golding, G. B. & Dean, A. M. Pervasive cryptic epistasis in molecular evolution. *PLoS Genet.* **6**, e1001162 (2010).
65. Natarajan, C. et al. Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* **340**, 1324–1327 (2013).
66. Doud, M. B., Ashenberg, O. & Bloom, J. D. Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol. Biol. Evol.* **32**, 2944–2960 (2015).
67. Starr, T. N. et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* **377**, 420–424 (2022).
68. Haddox, H. K., Dingens, A. S., Hilton, S. K., Overbaugh, J. & Bloom, J. D. Mapping mutational effects along the evolutionary landscape of HIV envelope. *eLife* **7**, e34420 (2018).
69. Viktorovskaya, O. V. et al. Divergent contributions of conserved active site residues to transcription by eukaryotic RNA polymerases I and II. *Cell Rep.* **4**, 974–984 (2013).
70. Mani, R., St Onge, R. P., Hartman, J. L. T., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. USA* **105**, 3461–3466 (2008).
71. Qiu, C. & Kaplan, C. D. Functional assays for transcription mechanisms in high-throughput. *Methods* **159–160**, 115–123 (2019).
72. Lin, X. et al. Nested epistasis enhancer networks for robust genome regulation. *Science* **377**, 1077–1085 (2022).
73. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
74. Sergey Ioffe, C. S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. in *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)* Vol. 37 448–456 (PMLR, Lille, France, 2015).
75. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
76. Malik, I., Qiu, C., Snively, T. & Kaplan, C. D. Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. *Nucleic Acids Res.* **45**, 4431–4451 (2017).
77. Scull, C. E., Ingram, Z. M., Lucius, A. L. & Schneider, D. A. A Novel Assay for RNA Polymerase I Transcription Elongation Sheds Light on the Evolutionary Divergence of Eukaryotic RNA Polymerases. *Biochemistry* **58**, 2116–2124 (2019).
78. Duan, B., Qiu, C., Lockless, S. W., Sze, S. H. & Kaplan, C. D. Higher-order epistasis within Pol II trigger loop haplotypes. *Genetics* **228**, iyae172 (2024).
79. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15**, 811–818 (2008).
80. Carvalho, A. T., Fernandes, P. A. & Ramos, M. J. The catalytic mechanism of RNA polymerase II. *J. Chem. Theory Comput.* **7**, 1177–1188 (2011).
81. Huang, X. et al. RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proc. Natl. Acad. Sci. USA* **107**, 15745–15750 (2010).
82. Castro, C. et al. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212–218 (2009).
83. Unarta, I. C., Goonetilleke, E. C., Wang, D. & Huang, X. Nucleotide addition and cleavage by RNA polymerase II: Coordination of two catalytic reactions using a single active site. *J. Biol. Chem.* **299**, 102844 (2023).
84. Mishanina, T. V., Palo, M. Z., Nayak, D., Mooney, R. A. & Landick, R. Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci. USA* **114**, E5103–E5112 (2017).
85. Palo, M. Z., Zhu, J., Mishanina, T. V. & Landick, R. Conserved Trigger Loop Histidine of RNA Polymerase II Functions as a Positional Catalyst Primarily through Steric Effects. *Biochemistry* **60**, 3323–3336 (2021).
86. Araya, C. L. et al. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci. USA* **109**, 16858–16863 (2012).
87. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* **19**, 1537–1551 (2013).
88. Harms, M. J. & Thornton, J. W. Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366 (2010).
89. Fowler, D. M. et al. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
90. Toulkikhonov, I., Zhang, J., Palangat, M. & Landick, R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* **27**, 406–419 (2007).
91. Yuzenkova, Y. et al. Stepwise mechanism for transcription fidelity. *BMC Biol.* **8**, 54 (2010).
92. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
93. Wang, W., Walmacq, C., Chong, J., Kashlev, M. & Wang, D. Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **115**, E2538–E2545 (2018).
94. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
95. Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).
96. Salinas, V. H. & Ranganathan, R. Coevolution-based inference of amino acid interactions underlying protein function. *eLife* **7**, e34300 (2018).
97. Cramer, P. Structure and function of RNA polymerase II. *Adv. Protein Chem.* **67**, 1–42 (2004).
98. Walmacq, C. et al. Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J. Biol. Chem.* **284**, 19601–19612 (2009).
99. Ziegler, L. M., Khapersky, D. A., Ammerman, M. L. & Ponticelli, A. S. Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF. *J. Biol. Chem.* **278**, 48950–48956 (2003).
100. Nouraini, S., Archambault, J. & Friesen, J. D. Rpo26p, a subunit common to yeast RNA polymerases, is essential for the assembly of RNA polymerases I and II and for the stability of the largest subunits of these enzymes. *Mol. Cell. Biol.* **16**, 5985–5996 (1996).
101. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).

102. Vos, S. M., Farnung, L., Linden, A., Urlaub, H. & Cramer, P. Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* **27**, 668–677 (2020).
103. Castro, C. et al. Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci. USA* **104**, 4267–4272 (2007).
104. Gregory, M. T., Gao, Y., Cui, Q. & Yang, W. Multiple deprotonation paths of the nucleophile 3'-OH in the DNA synthesis reaction. *Proc. Natl. Acad. Sci. USA* **118**, e2103990118 (2021).
105. Belogurov, G. A. & Artsimovitch, I. The mechanisms of substrate selection, catalysis, and translocation by the elongating RNA polymerase. *J. Mol. Biol.* **431**, 3975–4006 (2019).
106. Jenks, M. H., O'Rourke, T. W. & Reines, D. Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol. Cell. Biol.* **28**, 3883–3893 (2008).
107. Kuehner, J. N. & Brow, D. A. Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol. Cell* **31**, 201–211 (2008).
108. Qiu, C. et al. Thiolutin has complex effects in vivo but is a direct inhibitor of RNA polymerase II in vitro. *Nucleic Acids Res.* **52**, 2546–2564 (2024).
109. Greger, I. H. & Proudfoot, N. J. Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *EMBO J.* **17**, 4771–4779 (1998).
110. Kaplan, C. D., Holland, M. J. & Winston, F. Interaction between transcription elongation factors and mRNA 3'-end formation at the *Saccharomyces cerevisiae* GAL10-GAL7 locus. *J. Biol. Chem.* **280**, 913–922 (2005).
111. Simchen, G., Winston, F., Styles, C. A. & Fink, G. R. Ty-mediated gene expression of the LYS2 and HIS4 genes of *Saccharomyces cerevisiae* is controlled by the same SPT genes. *Proc. Natl. Acad. Sci. USA* **81**, 2431–2434 (1984).
112. Aguilera, A. Formamide sensitivity: a novel conditional phenotype in yeast. *Genetics* **136**, 87–91 (1994).
113. Sze, S. H. & Kaplan, C. D. Codon-Based Sequence Alignment for Mutation Analysis by High-Throughput Sequencing. In *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) 1-1* (Las Vegas, NV, USA, 2018).
114. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
115. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
116. Duan, B., Qiu, C., Sze, S. H. & Kaplan, C. Widespread epistasis shapes RNA Polymerase II active site function and evolution. (Zenodo, 2025).

## Acknowledgements

We thank Dr. Anne-Ruxandra Carvunis (U. Pittsburgh) and Dr. Steve Lockless (Texas A&M) for discussions and advice. We thank Zhizhen Wang and Muyao Lin from the Pitt Statistical Consulting Center for their

advice on checking the reproducibility of our data. We acknowledge funding from NIH R01GM097260 for initiation of this project and NIH R35GM144116 for this work. This research was supported in part by the University of Pittsburgh Center for Research Computing, RRID:SCR\_022735, through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483.

## Author contributions

B.D. performed sequencing library construction, data analysis, and made figures, drafted and revised the manuscript. C.Q. designed the mutant libraries, performed screening experiments, and contributed to revising the manuscript. S.H.Z. provided allele counts from raw sequencing data. C.D.K. conceived the project, guided analyses and interpretation of data, provided funding, and revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63304-6>.

**Correspondence** and requests for materials should be addressed to Craig Kaplan.

**Peer review information** *Nature Communications* thanks Alaksh Choudhury and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025