

**Title:** High-resolution phenotypic landscape of the RNA Polymerase II trigger loop

Chenxi Qiu<sup>1</sup>, Olivia C. Erinne<sup>3</sup>, Jui Dave<sup>1,5</sup>, Ping Cui<sup>1,6</sup>, Huiyan Jin<sup>1,7</sup>, Nandhini Muthukrishnan<sup>1,8</sup>, Leung K. Tang<sup>1,6</sup>, Sabareesh Ganesh Babu<sup>1</sup>, Kenny C. Lam<sup>1,9</sup>, Paul J. Vandeventer<sup>1,10</sup>, Ralf Strohner<sup>4</sup>, Jan Van den Brulle<sup>4</sup>, Sing-Hoi Sze<sup>1,2</sup>, Craig D. Kaplan<sup>1,11</sup>

<sup>1</sup>Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX

<sup>2</sup>Department of Computer Science and Engineering, Texas A&M University, College Station, TX

<sup>3</sup>Mississippi State University, Starkville, MS

<sup>4</sup>MorphoSys AG, Planegg, Germany

<sup>5</sup>Current address: Yale Cardiovascular Research Center, New Haven, Connecticut

<sup>6</sup>Current address: Department of Veterinary Pathobiology, Texas A&M University, College Station, TX

<sup>7</sup>Current address: Roche Nimblegen, Madison, WI

<sup>8</sup>Current address: Shire Pharmaceutical, Boston, MA

<sup>9</sup>Current address: College of Medicine, Texas A&M Health Science Center, College Station, TX

<sup>10</sup>Current address: Department of Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, TX

<sup>11</sup>Correspondence should be addressed to C.D.K. (cdkaplan@tamu.edu)

The active site of multicellular RNA polymerases have a “trigger loop” (TL) that multitasks in substrate selection, catalysis, and translocation. To dissect the *Saccharomyces cerevisiae* RNA polymerase II TL at individual-residue resolution, we quantitatively phenotyped nearly all TL single variants *en masse*. Three major mutant classes, revealed by phenotypes linked to transcription defects or various stresses, have distinct distributions among TL residues. We find that mutations disrupting an intra-TL hydrophobic pocket, proposed to provide a mechanism for substrate-triggered TL folding through destabilization of a catalytically inactive TL state, confer phenotypes consistent with pocket disruption and increased catalysis. Furthermore, allele-specific genetic interactions among TL and TL-proximal domain residues support the contribution of the funnel and bridge helices (BH) to TL dynamics. Our structural genetics approach incorporates structural and phenotypic data for high-resolution dissection of transcription mechanisms and their evolution, and is readily applicable to other essential yeast proteins.

RNA polymerase II (Pol II) synthesizes all eukaryotic mRNA. Structural studies of *Saccharomyces cerevisiae* (Sce) Pol II have illuminated mechanisms of transcription<sup>1-6</sup>, especially RNA synthesis. RNA synthesis occurs through iterative nucleotide addition cycles (NACs): selection of correct substrate nucleoside triphosphate (NTP), catalysis of phosphodiester bond formation, and enzyme translocation to the next template position. These critical steps in NAC appear to be coordinated by a critical, conserved domain within the Pol II active site: the trigger loop (TL).

TL functions are underpinned by its mobile and flexible nature (Fig. 1a). The primary function of TL is kinetic selection of correct NTP substrates while balancing transcription speed and fidelity, and this function is highly conserved based on studies of RNAPs from *E. coli* (Eco)<sup>7,8</sup>, *T. aquaticus* (Taq)<sup>9</sup>, the archaeon *P. furiosus* (Pfu)<sup>10</sup>, and eukaryotic Pol II from Sce<sup>11,12</sup> and human<sup>13</sup>. In a simplified two-step model, correct NTP binding appears to facilitate TL movement such that a bound, matched NTP shifts the TL from the “open” state to the “closed” state<sup>4,14-17</sup>, allowing capture of the matched NTP in the Pol II active site and promotion of phosphodiester bond formation<sup>4,16,18</sup>. The subsequent release of the byproduct, pyrophosphate, allows a conformational shift of the TL from the “closed” state back to the “open” state<sup>14,19,20</sup>. TL opening has been proposed to be critical for enzyme translocation relative to template, an essential step for the next nucleotide addition cycle<sup>8,12,14,21-24</sup>. Furthermore, additional TL states have been implicated in transcriptional pausing from studies in *E. coli*<sup>16,21,25</sup>, backtracking from structural observations<sup>26,27</sup>, and, although controversial, intrinsic cleavage<sup>7,28-31</sup>. Thus, distinct TL conformations or interactions are linked to different functions in transcription, with delicate control of TL dynamics promoting proper transcription elongation while possibly incorporating signals from the rest of Pol II or Pol II bound factors<sup>16,32-35</sup>.

Genetic and biochemical studies have revealed TL functions in the NAC. First, the nucleotide interacting region (NIR, Rpb1 1078-1085) discriminates matched rNTPs from 2'-dNTPs and non-complementary rNTPs<sup>11,12</sup>. NIR substitutions widely conferred lethality in residues observed to interact with rNTPs. Where viable, substitutions reduced catalytic activity *in vitro* and were termed as partially loss-of-function (LOF)<sup>7-11,36</sup>. Second, a TL C-terminal mutant E1103G, conferred increased catalytic activity *in vitro*, which we termed gain-of-function (GOF)<sup>11,12,37</sup>. Fast kinetics experiments revealed that E1103G may bias TL dynamics towards the catalytically active “closed” state<sup>12</sup>, consistent with infidelity and compromised translocation in addition to increased catalysis<sup>11,12,22,38,39</sup>. Furthermore, we previously described a set of Pol II TL mutants with broad and distinct alterations to transcription *in vivo*, thus conferring allele-specific phenotypes (Table 1) that correlate with decreased or increased activity<sup>36,40</sup> *in vitro*. Various genetic interactions (suppression, exacerbation, and epistasis) have also been observed, suggesting a complex functional network within the Pol II TL<sup>36</sup>. Finally, context dependence for TL function was observed, wherein analogous mutations in a conserved TL residue showed opposite effects in Sce Pol I and Pol II, suggesting different rate limiting steps for the two enzymes<sup>41</sup>. Together, the intricate intra-TL functional network and the context dependence of TL properties suggest importance of extensive residue-residue interactions within or outside the TL.

The possible multifunctional nature of each TL residue complicates interpretations of functions based on a limited number of mutants, since the phenotype of any given mutant could result from removal of the wild type side-chain or additional functions of the substituted one. Furthermore, different substitutions may have distinct effects on particular TL conformations<sup>36,42</sup>. In the TL, different substitutions in the same residue can

confer distinct phenotypes, so limiting mutational analyses to a single substitution at a particular position may mislead about residue function<sup>12,36</sup>. Deep mutational scanning is an emerging technique for studying large sets of mutants by assessing the enrichment or depletion of variants after a strict selection process<sup>43</sup>. Different selection approaches have been designed such that a specific protein property (sensitivity to substitutions<sup>44</sup>, thermo stability<sup>45</sup>, protein stability<sup>46</sup>, etc) can be studied. Notably, our established genetic phenotypes (Table 1) were well correlated with altered transcription elongation rates *in vitro* and multiple specific transcription defects *in vivo*<sup>36,40</sup>, thus providing a powerful phenotypic framework for studying TL function. In this work, we have defined the phenotypic landscape of the conserved, essential *S. cerevisiae* Pol II TL. We have found three major, distinct classes of transcriptionally defective TL mutants associated with differential stress response profiles, and determined the functional contributions of TL residues. We have interrogated the mechanisms by which proximal Pol II domains communicate with the TL, while identifying examples of epistasis between residues, which are the likely drivers of incompatibility of RNAP evolutionary variants when placed in the Pol II context.

## Results

### Strategy for studying *in vivo* effects of TL variants library

A comprehensively mutagenized TL variants library (Rpb1 1076-1106), with well-characterized TL mutants<sup>11,36</sup> excluded for further controls, was synthesized using the Slonomics technology<sup>47,48</sup> and validated by deep sequencing (Fig. 1b). Synthesis conditions were such that single substitution mutants would predominate. Our TL mutant library showed even distribution of substitutions across all positions and substitution types (Supplementary Fig. 1a-c), with generally very low frequencies for excluded mutants (Fig. 1b). We first sought evidence that measured allele frequencies reflected the real distribution of mutants, because PCR fidelity for similar amplicons is often compromised by template switching<sup>49,50</sup>. We spiked in five excluded single variants (H1085Y/Q, F1086S, G1097D, E1103G) as a control, so double mutants from the spike-in variants should not be present in our library but presumably the result of template switching. We prepared TL amplicons from a subset of conditions using both standard PCR and emulsion PCR (emPCR), which can suppress template switching<sup>49,50</sup>. First, double mutants from spike in controls were found at a significantly lower frequency than relevant single variants, and second, emPCR further suppressed the template switching frequency for all possible double mutants derived from spike-in single variants (Fig. 2a, left) 2.5 fold on average (Fig. 2a, right). We conclude that template switching is likely not extensive in our reactions but further reduction by emPCR led us to employ emPCR for our studies.

We have developed an experimental pipeline to examine mutations in an essential gene using a plasmid shuffling strategy, and have applied it to phenotyping the TL variant library (Fig. 1c). To validate our pipeline and to isolate novel TL alleles for further controls, we performed a traditional genetic screening for mutants with transcriptional defects (Table 1) and isolated 1166 candidate mutants (Supplementary Table 1), which included 154 singly-substituted and 386 multiply-substituted variants. To further distinguish mutants, we examined 50 single variants under various stress conditions (Fig. 2b, Supplementary Fig. 2a), and observed that media containing caffeine, hydroxyurea, MnCl<sub>2</sub>, formamide, cycloheximide, or NaOH induced allele-specific sensitivity or resistance, while media containing ethanol, benomyl, HCl or NaCl showed fewer allele-specific effects (Fig. 2b, Supplementary Fig. 2a). Therefore, in our high-throughput approach, we phenotyped TL variants library under our established

conditions (Spt<sup>-</sup>, MPA<sup>S</sup>, Gal<sup>R</sup>) and stress conditions exhibiting allele-specific discrimination ability. Phenotypic score was estimated from the change of allele frequency normalized to WT, as is standard for previous mutational scanning studies<sup>43-46</sup>. Quantitative phenotypic scores of the 50 mutants from the high-throughput phenotyping were consistent with semi-quantitative growth scores derived from standard phenotyping (Fig. 2c, supplementary Fig. 2), validating our approach.

### The Pol II TL fitness landscape

The TL is highly conserved, especially in the NIR, the loop tip residue (Rpb1 G1088) and several TL C-terminal residues (Fig. 3a). Highly conserved residues are predicted to be critical for protein function, thus substitutions are expected to confer fitness defects and be selected against. We first sought to evaluate general fitness defects of observed TL singly-substituted variants (termed “fitness landscape”), both in the presence of WT *RPB1* (Fig. 3b) and upon removal of WT *RPB1* (Fig. 3c). Notably, TL NIR and loop tip substitutions conferred large fitness defects in general, while most perturbations in the similarly conserved C-terminal residues did not confer severe growth defects (Fig. 3b, 3c). This observation highlights that conservation does not necessarily reflect sensitivity to perturbations, and TL fitness landscape can further distinguish extremely highly conserved TL residues, as discussed below:

First, substitutions in the NIR (Rpb1 1077-1085) generally conferred both fitness defects (Fig. 3c) and apparent dominance (Fig. 3b). Observed fitness defects were consistent with previous observations that several NIR mutants render Pol II slow in elongation *in vitro* and cause fitness defect *in vivo*<sup>11,36</sup>. Slowly elongating Pol II may engender clashes with WT Pol II on genes *in vivo*, consistent with the observed dominance in most NIR variants, and consistent with TL variants being assembled into Pol II complexes to interfere with WT Pol II function. Second, substitutions within the alanine-glycine linker (Rpb1 1087-1088) almost universally conferred lethality or severe growth defects. A Pol II structure with closed TL<sup>4</sup> reveals that A1087 and G1088 are in a tight pocket between the funnel and bridge helices, presumably necessitating small side-chain residues (Supplementary Fig. 3a). To determine the extent of spatial constraint, we individually assessed the fitness of AG swapping variants and small hydrophobic valine substitutions (Fig. 3d). Notably, all the swapping variants (A1087G, G1088A and A1087G/G1088A) were lethal (Fig. 3d). While G1088V is lethal, A1087V is severely sick but viable (Fig. 3d), suggesting extremely high spatial constraint but differential tolerability for the two residues. This pocket/TL interaction is only observed in the closed TL<sup>4</sup> but not in any of the open states<sup>51</sup>, suggesting function in stabilizing the active, closed TL conformation for promoting catalysis. Consistent with disruption of the pocket/TL interaction and the closed TL state, we observed LOF phenotypes for A1087V (Gal<sup>R</sup>, slight MPA<sup>R</sup>) (Fig. 3e). Finally, substitutions in the conserved C-terminal helix, though not strongly defective in general fitness, might have transcription defects and were further characterized (discussed below).

### Novel TL NIR mutants allow mechanistic insights

The TL fitness landscape identified residues highly sensitive to perturbations, while also revealing variants in NIR residues previously known to be difficult to viably substitute. We highlight L1081 and H1085 as two examples. L1081 directly interacts with the base of matched NTPs<sup>4</sup>, and the equivalent residues in *Eco* and *Taq* RNAPs are important for substrate selection or catalysis<sup>7,9</sup>. L1081 is the most sensitive to perturbations among the hyper-conserved NIR. All previously tested L1081 variants were lethal<sup>36</sup>, though viable substitutions were identified for all other NIR residues. Furthermore, GOF allele



E1103G can generally suppress lethal substitutions for most NIR residues, but could not for tested L1081 substitutions. In our TL fitness landscape, almost all L1081 variants were indeed predicted to be lethal based on our fitness threshold (Fig. 3c). L1081M conferred a severe growth defect, but was predicted to be just above the viable threshold (Fig. 3c). To validate this prediction, we constructed L1081M for direct analysis, and found that L1081M was indeed viable yet severely sick (Fig. 3d). Furthermore, L1081M conferred Gal<sup>R</sup> and slight MPA<sup>R</sup> phenotypes, consistent with other LOF mutants (Fig. 3e). Eukaryotic multi-subunit RNA Polymerases share a stringent evolutionary requirement for L at this TL position, while bacterial and archaeal lineages show both M and L variants. Consistent with evolutionary tolerance of variation within bacterial and archaeal lineages, the *Taq* RNAP M1238L variant shows near WT activity for substrate selection and catalysis *in vitro*<sup>9</sup>. These results highlight epistasis within *Sce* Pol II, and likely eukaryotic RNAP lineages, imposing a stringent requirement for Leucine.

H1085 interacts with the  $\beta$ -phosphate of the matched NTP<sup>4</sup>, and has been implicated in substrate selection, catalysis, intrinsic cleavage and PPI release<sup>28,52</sup>. We previously constructed several H1085 variants (A/N/D/F were lethal, K/R/W/Y caused severe growth defects, Q caused slight growth defect<sup>11,36,40</sup>), suggesting that some polar or positively charged residues, but not a hydrophobic phenylalanine or alanine, could partially complement loss of the histidine<sup>36</sup>. Here, we found that H1085L was viable and healthy in the fitness landscape (Fig. 3c), and validated it with individual mutant analysis (Fig. 3d). While H1085L conferred slight MPA<sup>R</sup> and Gal<sup>R</sup> phenotypes, consistent with other LOF mutants (Fig. 3e), it also conferred a slight Spt<sup>-</sup> defect, suggesting distinct defects from most other NIR mutants and all known LOF mutants. This observation alters our understanding of the likely bounds of active site chemistry (see discussion).

### **There are at least three distinguishable TL mutant classes**

The overall TL fitness landscape revealed the essentiality of almost all single substitution TL variants in standard growth medium, but could not indicate the nature of transcriptional defects, as we had previously found that both LOF and GOF alleles conferred growth defects. Therefore, we sought to determine the phenotypic outcome of the TL variants for the transcription-related Gal<sup>R</sup>, MPA<sup>S</sup> and Spt<sup>-</sup> phenotypes as well as for a variety of stress conditions (termed “phenotypic landscape”).

Hierarchical clustering of the phenotypic landscape for 412 TL variants passing fitness filters revealed three mutant classes with distinct features (Fig. 4a, 4b). Class 1 mutants generally conferred strong Gal<sup>R</sup> phenotype yet were Spt<sup>+</sup>, and in some cases were also slightly MPA<sup>R</sup> relative to WT, consistent with previously characterized LOF mutants. We also identified high formamide sensitivity as a new signature phenotype for Class 1 mutants. Class 2 mutants showed generally weaker Gal<sup>R</sup>, slight formamide resistance, but do not confer strong phenotypes otherwise, representing a novel TL mutant class yet to be biochemically characterized. Class 3 mutants generally conferred Gal<sup>R</sup>, Spt<sup>-</sup> and MPA<sup>S</sup> phenotypes, consistent with previously characterized GOF mutants. Mn<sup>2+</sup> hypersensitivity (Mn<sup>S</sup>) was correlated broadly with Spt<sup>-</sup> and MPA<sup>S</sup> phenotypes, suggesting a relationship among these phenotypes, and consistent with previous *in vitro* biochemical and *in vivo* phenotypic data for a subset of known GOF mutants<sup>53,54</sup>. Notably, our spike-in LOF (F1086S, H1085Q and H1085Y) and GOF mutants (E1103G and G1097D) co-clustered with Class 1 and Class 3 mutants, respectively.

## Functional contribution of TL residues in different states and substrate-induced TL closing mechanism

The distributions within different mutant classes predict distinct functional contributions of TL residues to TL dynamics. Perturbations predicted to bias the TL towards the active, closed TL state have been shown to result in GOF, whereas destabilization of the closed TL state generally leads to LOF<sup>8,11,12,16,36</sup>. Therefore, distributions of Class 1 (LOF) and Class 3 (GOF) mutants predict alterations to TL dynamics, as follows:

Class 1 (LOF) mutants included most variants from F1086, V1089, V1094 and P1099 (Fig. 4c, left), suggesting functions of these residues in stabilizing the closed TL. F1086 and V1089 are both proximal to multiple funnel helix residues when TL is closed<sup>4,17</sup>, while F1086 was proposed to orient H1085 for correct substrate interaction<sup>17</sup>. Therefore, alteration of these interactions may disrupt the closed TL state and result in LOF. Alternatively, recent Pol II structures revealed potential function of F1086-V1089 interaction in TL closing dynamics (Supplementary Fig. 3b). V1089 forms a backbone-backbone hydrogen bond with F1086 when TL is open, while its side chain flips towards the F1086 to form a hydrophobic interaction when TL is partially closed, suggesting that this side-chain interaction may be important for particular TL states (Supplementary Fig. 3b), though it was not discussed in previous molecular dynamics (MD) studies<sup>17</sup>. Furthermore, V1094 was observed to be proximal to the BH residue K830 in the closed TL state<sup>4</sup>. An interaction between K830 and V1094 side-chains could be counter-intuitive and possibly undervalued. However, neutralization of lysine's positive charge through ionic interactions (such as D836) can promote hydrophobicity of the lysine side chain<sup>55</sup>, supporting the observed K830-V1094 interactions in the TL closed state (Supplementary Fig. 3c). Most variants in V1094 are LOF (Fig. 4c, left), consistent with disruption of K830-V1094 interaction and concomitant destabilization of the closed, active TL conformation.

Models for NTP substrate-induced TL closing remain largely untested<sup>4,14-17</sup>. A recent Pol II structure<sup>51</sup> exhibiting an open TL state led to explicit implication of a hydrophobic pocket formed by TL residues (A1076, M1079, T1080, G1097 and L1101) and other TL proximal residues (I837, L841, V1352, V1355 and I1356) in substrate-induced TL-folding (Supplementary Fig. 3d). Q1078 recognition of the 2'-OH of a matched NTP substrate was proposed to promote release of the adjacent residue M1079 from the hydrophobic pocket, triggering TL closing<sup>51,56</sup>. Consistent with disruption of this observed pocket and concomitant destabilization of the inactive open TL state, A1076T, a pocket variant previously isolated as genetically GOF, conferred increased transcription activity *in vitro* (Fig. 5b). Notably, GOF phenotypes were observed for a large number of variants in pocket residues. Among them, we observed almost universal GOF phenotypes for G1097 variants, but not the extreme fitness defects found for the previously observed GOF variant G1097D. We individually phenotyped ten G1097 variants from the traditional screening and confirmed this observation (Supplementary Fig. 3e). Together, these results are consistent with the hydrophobic pocket stabilizing the inactive, open TL and providing a plausible mechanism for substrate-induced TL closing where a single residue, M1079, can act as linchpin for the entire TL through a network of interactions.

## Identification of stress conditions that alter transcription *in vivo*

GOF and LOF TL variant classes have distinct stress response profiles. In general, compared to LOF variants, GOF mutants are more sensitive to Mn<sup>2+</sup>, caffeine and cycloheximide yet more resistant to hydroxyurea and formamide (Fig. 4d). The allele-specific Mn<sup>2+</sup> response amplified our previous observation that GOF E1103G was highly

sensitive to  $Mn^{2+}$  while the LOF H1085Y was resistant to  $Mn^{2+}$ , and  $Mn^{2+}$  effects on both mutants were suppressed by  $Mg^{2+}$  supplementation<sup>52</sup>. The TL phenotypic landscape showed that this  $Mn^{2+}$  response was general and class-specific for GOF and LOF mutants (Fig. 4d). To validate this observation, we individually analyzed seven additional variants (two LOF and five GOF) for  $Mn^{2+}$  sensitivity in the presence or absence of  $Mg^{2+}$  supplementation. Notably, all tested LOF mutants conferred  $Mn^{2+}$  resistance while all tested GOF mutants conferred  $Mn^{2+}$  hypersensitivity (Fig. 4e). Allele-specific  $Mn^{2+}$  responses could be suppressed by  $Mg^{2+}$  supplementation (Fig. 4e).  $Mn^{2+}$  has been shown to stimulate transcriptional activity while compromising fidelity *in vitro*<sup>53,54</sup>. Our observation suggested the possibility that  $Mn^{2+}$  may suppress LOF mutants by stimulating transcriptional activity yet exacerbate GOF mutants by further decreasing their already compromised transcriptional fidelity *in vivo*<sup>11,12</sup>. Increased Pol II catalytic activity correlates strongly with upstream transcription start site (TSS) shift *in vivo*<sup>36,40</sup>, therefore we assayed for TSS alterations upon  $Mn^{2+}$  treatment. Primer extension analysis at *ADH1* revealed that  $Mn^{2+}$  treatment shifted the TSS distribution upstream, and further exacerbated the upstream shift in E1103G (Fig. 4f). Deletion of *PMR1*, the golgi  $Mn^{2+}$  export channel, causes accumulation cytosolic  $Mn^{2+}$ <sup>57,58</sup>, and can be used to alter  $Mn^{2+}$  levels apart from supplementing medium. Our prior high throughput genetic interaction analyses of Pol II mutants showed that *pmr1Δ* strongly interacts with Pol II mutants in a highly allele-specific fashion, suggesting an intimate relationship between increased cellular  $Mn^{2+}$  levels and altered transcription activity. Here we find that *pmr1Δ* also shifted *ADH1* TSSs upstream (Fig. 4f). While  $Mn^{2+}$  may have other indirect effects on Pol II mutants, these observations support direct effects of  $Mn^{2+}$  on Pol II transcription activity *in vivo*, raising the possibility that other allele-specific stress conditions (e.g. formamide) may also directly alter transcription *in vivo*.

### Functional contributions of the TL tip region

The TL tip region (Rpb1 1090-1096) is a random-coil region that forms an  $\alpha$ -helical structure when the TL is closed, and helical formation has been proposed to assist TL closing<sup>8,17,42</sup>. Mejia *et al* characterized two *Eco* RNAP TL tip mutants I1134V and G1136S (Equivalent to *Sce* Pol II V1094 and S1096) with slightly decreased or increased transcription activity, respectively<sup>42</sup>. These results were interpreted as I1134V and G1136S substitutions decreasing or increasing the helical propensity and thus disfavoring or favoring TL closing<sup>42</sup>. *Sce* Pol II contains each of these variants as the WT residue, therefore individual substitutions to the *E. coli* variants (V1094I and S1096G) would be predicted to confer opposite phenotypes under the helical propensity model. However, V1094I and S1096G did not confer clear phenotypes consistent with either GOF or LOF (Fig. 4b), failing to support the helical propensity model. We asked if the proposed correlation from *Eco* RNAP studies was a general property for TL substitutions in this region, if extended to more than two substitutions. Our data, calculated from 122 variants, fail to support a general correlation between helical propensity and predicted catalytic activity for Pol II substitutions in this region (Supplementary Fig. 6a). As discussed above, V1094 may be involved in interaction with BH residue K830, and LOF in most V1094 variants may result from disrupted BH/TL coordination. Therefore, we repeated the analyses excluding V1094 variants, yet still failed to observe a correlation (Supplementary Fig. 6a). We cannot rule out contributions of helical propensity in this region to TL function; however, we did not find compelling or widespread evidence for it.

A number of recent studies have suggested potential functions of the TL tip region in regulating TL dynamics<sup>17,51,59</sup>. In a simulated TL closing process, positively charged K1092 and K1093 were predicted to interact with several TL-proximal residues, and

some of the predicted interactions were validated by subsequent Pol II crystal structures with alternative open TL states (Fig. 5a). These interactions were proposed to stabilize the open, inactive TL state, and thus alanine (K1092A, K1093A) or charge reversing substitutions (K1092D/E, K1093D/E) were predicted to disrupt the inactive TL open state and result in GOF<sup>17</sup>. Contrary to this prediction, none of the above substitutions conferred GOF (Fig. 4b). Networks of residue-residue interactions near the TL tip were observed<sup>17,51</sup>, some of which may be functionally overlapping or redundant, adding complexity to simple models. Our previous point mutant epistatic miniarray profile (p-EMAP) studies predicted two TL-proximal mutants (S713P and I1327V) to be GOF, which we confirm here (Fig. 5b), suggesting that perturbation near the TL may interfere with native interactions, or create new ones, to destabilize the open TL. The tested variants here also extend the correlation between genetically predicted GOF and increased activity *in vitro* (Fig. 5b). Additionally, several TL tip variants with bulky side chains (K1092W, K1093Y, K1093M) conferred GOF phenotypes (Fig. 4b). Given the complexity and observation of both GOF/LOF phenotypes, we wished to further assess the functions of these residue-residue interactions.

Functional interactions among residues can be explored by the similarity between single substitution variants and the phenotypes of double mutants. We first sought evidence that variants in the potential interaction partners could confer similar GOF or LOF phenotypes. In the simulation, K1092 switched interaction partners between two funnel helix residues D716 and E712<sup>17</sup>, and other charged residues were either observed or simulated to interact with S1091, K1092 or K1093 (Fig. 5a). Therefore, we constructed a panel of mutants in the residues D716, E712, R1281, E1307, and D1309 for phenotypic analyses. Notably, we observed GOF phenotypes (Mn<sup>S</sup> and MPA<sup>S</sup>) in E1307K but not E1307A, suggesting that E1307K gained an interfering interaction to destabilize the open TL state. Furthermore, we observed the Gal<sup>R</sup> phenotype in D716A (Fig. 5d, Supplementary Fig. 6f), consistent with LOF. D716K and E712A were lethal (Fig. 5d, Supplementary 6b, 6c), but their defects were further explored by double mutant analyses (discussed below). Together, both GOF and LOF variants were observed in the TL tip proximal residues, consistent with roles in regulating TL dynamics.

To further dissect functional relationships, we phenotyped double mutants from potential interaction partners, and observed a number of genetic interactions (Fig. 5d, Supplementary Figure 6g-k). First, GOF and LOF mutants were mutually suppressive when combined, and most TL mutants from same biochemical class (GOF/GOF or LOF/LOF) showed additive effects (synthetically sick or lethal). The observed class-specific genetic interactions are similar to the previously reported intra-TL genetic interactions<sup>36</sup>, consistent with alteration of TL function in TL tip proximal variants. Furthermore, K1092A/D single substitutions did not confer transcription phenotypes, but were able to suppress the E1307K GOF phenotypes. This observed epistasis suggested that loss of K1092 relieved a putative gain of interaction in E1307K (discussed above). Finally, E712A lethality was fully suppressed by K1092A, K1092D or K1093M, adding an additional instance of epistasis. A model to explain this complex genetic relationship is that loss of native E712-K1092 interaction re-directed K1092 towards an alternative interaction or strengthened an existing interaction with D716, causing lethality. Alteration of TL tip interaction potential through K1092/1093 substitutions relieves this allele-specific effect. Taken together, the observed allele-specific and epistatic interactions between TL tip and proximal residues suggest a highly complex genetic network of residues controlling TL dynamics, and illustrate how individual residues might constrain or allow diversification of the TL through evolution.



## Functional interplay of TL and Bridge helix (BH)

The BH is a strikingly conserved structural domain of multisubunit RNA polymerases spanning the wide central cleft between polymerase “jaws”, adjacent to the active site and proximal to the TL<sup>1,60,61</sup>. Although BH is a straight helix in most published structures<sup>1-6</sup>, some *Thermus thermophilus* RNAP structures reveal a bent BH conformation proposed to support translocation<sup>60</sup>. This BH bending mechanism was supported by a number of simulation studies but has never been directly tested<sup>1,24,60-62</sup>. In the archaea *Methanocaldococcus jannaschii* (*Mja*) RNAP, proline substitutions at two hinge-proximal residues M808 and S824 (equivalent to *Sce* Rpb1 M818 and T834) resulted in GOF, suggesting kinking by proline substitution results in increased translocation or catalysis<sup>62,63</sup>. Furthermore, *Mja* GOF TL and BH mutants were not additive when combined, suggesting mutual dependence on BH and TL functions<sup>62</sup>.

To explore the functional consequence of BH kinking in *Sce* Pol II, we constructed and phenotyped BH mutants analogous to the characterized GOF and LOF variants in *Mja* RNAP. Notably, *Sce* T834 and other BH C-terminal hinge substitutions conferred *in vivo* phenotypes consistent with the altered transcriptional activities in *Mja* RNAP (Fig. 6a), and we directly confirmed the altered activity of T834 variants *in vitro* (Fig. 6b). In contrast, substitutions in M818, a predicted BH N-terminal hinge, showed defects deviating from expected conservation of function. M818P caused lethality, and could not be suppressed by all tested TL variants, precluding us from classifying it (Supplementary Fig. 7a). Furthermore, M818S and M818Y, although viable, did not confer any clear phenotypes (Fig. 6a). Therefore, we further assessed the functional interplay between BH and TL by double mutant analyses, including BH variants (M818S/Y, T834A/P) and TL substitutions covering a range of altered transcriptional activities (Fig. 6f-i). Notably, the GOF BH variant T834P, along with M818S and M818Y, were mutually suppressive with biochemically strong LOF TL variants (Fig. 6c, 6d, 6f), revealing both additive behavior between BH and TL for some combinations, and cryptic phenotypes for M818S/Y. The LOF BH variant T834A also suppressed GOF TL variants (Fig. 6e). However, the additive interactions (exacerbation, synthetic lethality) we observed for GOF BH and TL double mutants were in contrast to the epistasis for *Mja* RNAP<sup>62</sup>.

Multiple lines of evidence suggested additional, specific defects exist in BH mutants, beyond simple cooperation with the TL. First, M818P lethality could not be suppressed by all tested TL variants (Supplementary Fig. 7a), which cover a wide range of transcriptional activities. Second, suppression between BH and TL mutants of different biochemical classes (GOF/LOF) was partial and not as strong as previously observed intra-TL suppression. Third, GOF M818S, M818Y and T834P variants appeared to exhibit activity-dependent genetic interactions with TL variants. BH GOF variants suppressed strong LOF TL variants Q1078S and H1085Y but failed to suppress, or even exacerbated slightly LOF TL variants H1085Q and F1086S (Fig. 6c, 6d, 6f), consistent with conditional epistasis, where GOF activity of BH variants can suppress either specific TL variants or otherwise exert their effects in specific contexts. Finally, recent modeling studies predicted BH residue Y836 to assist Pol II forward translocation<sup>64</sup> by interacting with the DNA:RNA hybrid. Y836A/H conferred Gal<sup>R</sup> phenotypes, consistent with LOF and compromised translocation (Fig. 6a). Notably, GOF T834P was suppressed by Y836A/H (Fig. 6f), consistent with T834P conferring TL independent fast translocation defect, suppressible by Y836A/H.

## Context dependence of TL function

We previously observed that E1103G, a GOF allele in *Sce* Pol II, caused LOF in Pol I, highlighting divergent contributions of active site residues in different enzymatic contexts<sup>41</sup>. We also observed that Pol I TL and L1081M (in this study) were functionally impaired in the Pol II context. We next sought to determine the functional compatibility of other evolutionary TL variants in the *Sce* Pol II context, using our fitness and phenotypic landscape (Fig. 7). Most tested evolutionary TL variants did not confer fitness defects, though several did (Fig. 7a). Furthermore, some variants, although compatible for general growth, conferred transcription-related phenotypes and could be further classified by phenotypic landscape (Fig. 7b). These observations further demonstrate that the evolution of TL function is shaped by likely epistasis between the TL and proximal domains.

We next asked what substitutions might underlie the large difference in compatibility of the *Sce* Pol I TL (versus the *Sce* Pol III TL) within Pol II<sup>41</sup>. From our phenotypic landscape, although many individual *Sce* Pol I and Pol III TL substitutions appeared to be compatible, functionally impairing variants were identified (Fig. 7b). The yeast Pol III TL contains Pol II GOF (A1076G) and LOF (N1082K) variants, both of which hypothetically could be mutually suppressive, resulting in close to WT activity in the Pol II context<sup>41</sup>. The Pol I TL contains three Pol II LOF mutations (V1089H, A1090G and S1091A). The net incompatibility of Pol I TL is consistent with additive defects of the three LOF variations, given that most TL LOF combinations show additive effects<sup>36</sup>. Since three evolutionarily observed variants with LOF phenotypes were all localized in the TL tip, we examined the difference between Pol I and Pol II structures for the TL tip proximal domains<sup>51,65</sup>. The Pol I funnel helix appears to impose less constraint than the Pol II funnel helix (Fig. 7c), suggesting that Pol I controls its TL with a distinct network of interactions. In all, our mutational data, together with the recent Pol I crystal structure, reveals enzyme-specific mechanisms to control a highly conserved domain at the heart of eukaryotic transcription.

## Discussion

The ability of the TL to fold into multiple conformations and the dynamic conversion between these states are critical for its functions. Previous studies from us and others demonstrate that TL function is delicately balanced, such that perturbations result in either increased or decreased catalytic activity and altered translocation dynamics. Distinct consequences for transcriptional activity manifest *in vivo* as what we term LOF and GOF phenotypes. In this study, we have advanced our genetic framework with which to dissect Pol II mechanisms. From our phenotypic landscape, we assessed the functional contributions of almost all TL residues to fitness in *S. cerevisiae* under multiple conditions. Our data indicate that both intra-TL interactions and TL interactions with nearby domains (e.g. funnel helix) are critical for TL function. This conclusion is also supported by recent work on Rpb9 organizing the TL indirectly through an Rpb1 TL-adjacent loop<sup>59</sup>, interactions between the TL and F-loop regions in bacteria<sup>30</sup>, and predictions of TL proximal variants as GOF from our previous pEMAP analysis<sup>40</sup> (validated in this study). Our system allows efficient analysis of a large number of variants to evaluate accumulating computational<sup>17</sup> and structural<sup>51</sup> predictions for interactions within the TL and from without.

The major function of the TL is to link substrate recognition to catalysis, and is also proposed to gate translocation such that translocation probability is linked to phosphodiester bond formation. Critical to this recognition is that a substrate be positioned correctly by base-pairing to the DNA template, and that the 2'-OH allows

NTPs to be selected over 2'-dNTPs by the TL residue Q1078<sup>4,9,27</sup>. We have proposed that the Q1078-substrate interaction releases the adjacent M1079 from its intra-TL hydrophobic pocket, and to trigger the TL closing<sup>51</sup>. In this study, we show a great number of variants from the pocket residues A1076, M1079, G1097, L1101 causing GOF, providing evidence that disruption of the hydrophobic pocket destabilizes the open, inactive TL state, resulting in GOF. Additionally, while the TL shows incredibly high evolutionary conservation for a number of residues, prior work indicated alteration of ultra-conserved residues in different RNA polymerases could have distinct effects, suggesting the importance of the evolved context within each enzyme<sup>11,36,41</sup>. Here, we evaluate many evolutionarily observed eukaryotic TL variants in the *Sce* Pol II system, and discover a number of functionally impaired TL variants. Our results highlight that TL proximal domains may impose constraint and possibly also allow functional diversification in the molecular evolution of highly conserved TL by epistatic interactions.

The characterization of the unexpectedly healthy H1085L variant clouds the issue of how H1085 functions in substrate selection and catalysis. H1085 interacts with the substrate NTP through salt bridge and hydrogen bond<sup>4</sup>, and previous simulations with limiting H1085 variants predicted the hydrogen bonding to be critical for maintaining substrate interaction<sup>66</sup>. The discovery of H1085L argues that productive substrate interactions may be supported by entirely different chemistry, although we cannot rule out the possibility that H1085L redirects substrate interactions to an alternative residue. Furthermore, H1085 variants may have multiple defects in NAC, such as substrate selection<sup>11</sup>, catalysis<sup>11,52</sup>, intrinsic cleavage<sup>52</sup> and PPi release<sup>19,20</sup>, and whether or not H1085 or analogous residues act as a general acid remains controversial in different RNAPs<sup>4,7,9,52,67</sup>. Function of H1085L in all of these steps remains to be determined, but its phenotype suggests that function of H1085 as a general acid may be entirely bypassed.

The established TL phenotypic landscape can be further explored to study intra- and inter TL epistasis. First, whether individual TL residues work collaboratively or independently to ensure balanced TL dynamics and proper function is an open question. Some TL residues can be functionally overlapping acting at similar steps, or functionally discrete acting at distinct steps. For example, combination of LOF mutations in Q1078, N1082 and a TL-proximal residue N479 resulted in non-additive genetic interaction, suggesting functionally overlapping roles of these residues. In contrast, combination of variants from Q1078 (or N1082) and H1085 resulted in exacerbation or synthetic lethality, suggesting independent functions<sup>36</sup>. Coupled with structures of partially folded TL states, these genetic studies support the functional distinction between NIR residues and a multi-step TL folding model for the promotion of catalysis<sup>36</sup>. Here, we have identified many more predicted GOF and LOF TL variants (Fig. 4b), some of which are predicted to confer epistatic interactions (e.g. F1086 and V1089). We expect the phenotypic landscape of a multiply-substituted TL library to be extremely informative for understanding functional relationship between TL residues.

Second, the TL phenotypic landscape is an extremely sensitive readout for assessing active site re-arrangement. Transcription is under control by many factors, some of which may alter the Pol II active site conformations, though few studies directly address these possibilities. Initiation factors and Pol II TL mutants confer similar alterations in transcription start site selection, consistent with initiation factors functioning through the Pol II active site and altering the efficiency of Pol II catalysis during initiation<sup>40,52,68</sup>. Furthermore, TL may communicate with other Pol II sites, such as the RNA exit channel

or clamp domain<sup>35</sup>, or in direct competition with external factors, such as TFIIIS<sup>32</sup>. Perturbations of this communication may alter TL dynamics and cause allele-specific genetic interactions (Fig. 5, Fig. 6). Specifically, an external perturbation by a relevant factor or Pol II TL distant domain may show epistasis or synergy only with specific TL alleles of a class (either LOF or GOF), whereas a non-interacting factor may not. Finally, similar perturbation of the TL phenotypic landscape by different factors would suggest functional similarity between them, thus clustering of phenotypic landscape changes upon different perturbations is expected to provide valuable insight.

The TL phenotypic landscape, along with our previous work<sup>36</sup>, illustrates a strategy of utilizing *in vivo* genetic reporters or stress response profiles to distinguish a large number of mutants with distinct *in vivo* defects. As discussed above, the phenotypic landscape sheds light on functional contribution of TL residues to its dynamics, to the mechanism of catalysis and to the evolutionary constraints of the TL sequence and function. The phenotypic landscape strategy expands the current scope of existing deep mutational scanning studies<sup>43-46</sup>, and can be generalized to study most, if not all, yeast proteins.

## Acknowledgments

We gratefully acknowledge Stefan Heiderich for technical assistance during the Slonomics library synthesis. We thank members of Kaplan lab, Steve Lockless, Qiuyan Shao, and Kun Ping Lu for helpful discussion and feedback on the manuscript. This work was supported by National Institutes of Health grant R01GM097260 and Welch Foundation Grant A-1763 to CDK, and by National Institutes of Health Grant R01GM036659 to Roger Kornberg during the initial stages of this project. We also acknowledge NSF REU site grant DBI-0851611 for supporting O.E.



## Online Methods

### Yeast strains, Media and Plasmids

All yeast strains are derived from a *GAL2+* derivative of S288C strain<sup>69</sup>. Genotypes and phenotypes of the yeast strains are in [Supplementary file 1](#). Standard yeast media and the media for assessing established transcription phenotypes are as described previously<sup>36</sup>. For studies with 15mM Caffeine (Sigma), 150mM Hydroxyurea (Sigma), 5mM and 15mM  $MnCl_2$  (Sigma), 0.5M NaCl (EMD), 3% formamide (JT Baker), 6% ethanol (KOPTec), 0.07 $\mu$ g/mL cycloheximide (Sigma), 10mM HCl (EMD), 10mM NaOH (EMD), 10 $\mu$ g/mL Benomyl (Sigma), each compound was added to the minimal SC-Leucine medium at the indicated concentration from concentrated stock solutions.

Detailed description of plasmids is in [Supplementary file 1](#), and complete sequences of plasmids are available upon request. For studies involving individual analyses of Pol II mutants, site-directed mutagenesis was performed via the Quikchange strategy from Stratagene. All mutagenized regions have been verified by sequencing before sub-cloning into pRS315-derived plasmids, as previously described<sup>36</sup>.

### Genetic and biochemical analyses of individual Pol II mutants

Phenotypic analyses of individual Pol II mutants were performed by plasmid shuffling assays, with viable mutants further subjected to standard plate phenotyping. Each mutant in a pRS315-derived plasmid (*CEN LEU2*) was transformed into CKY283 (*rpb1 $\Delta$ ::CLONATMX*, pRP112 *RPB1 CEN URA3*). Transformants (Leu<sup>+</sup>) were patched on SC-LEU plates and subsequently replica plated to SC-LEU+5FOA (1mg/mL) to assay complementation ability upon loss of the *RPB1 CEN URA3* plasmid. Experimental details are as previously described<sup>11,36</sup>. Saturated cultures from single colonies of viable and shuffled Pol II mutants were subject to 10-fold serial dilution and spotting on indicated phenotyping media, as described in various previous reports<sup>11,36</sup>.

Pol II enzymes were purified via a tandem-affinity tag (TAP) protocol derived from<sup>70</sup> with modifications described in<sup>11</sup>. Transcription elongation reactions were performed with Pol II elongation complexes assembled on a nucleic acid scaffold, in a procedure described in<sup>11</sup> with slight modifications in the amount of Pol II and nucleic acids as described in<sup>51</sup>. For each enzyme, elongation assays were performed with 25 $\mu$ M, 125 $\mu$ M, 500 $\mu$ M and 750 $\mu$ M NTPs (each of ATP, GTP, CTP, UTP), and maximal elongation rates were extracted exactly as previously described<sup>11</sup>.

*ADH1* transcription start site selection was analyzed by primer extension. In brief, indicated strains were grown in YPD until mid-log phase ( $\sim 1 \times 10^7$  cells/mL), and diluted with YPD with 10mM  $MnCl_2$  or equal volume of  $H_2O$ . Total RNA was extracted as described<sup>71</sup>, and 30 $\mu$ g of total RNA was subject to primer extension analysis, following a protocol derived from<sup>72</sup> with modifications described in<sup>36</sup>.

### High-throughput phenotypic analyses of the TL variants library

The TL variant library was synthesized by Sloning Biotechnology (now MorphoSys) with well-characterized TL variants excluded (specified in Fig. 1B) using a building block approach<sup>47,48</sup>. The TL variant library was transformed into CKY283 via a gap-repair strategy as previously described<sup>40</sup>. In brief, the amplified TL variant library with flanking sequence was transformed into CKY283 together with a linearized pRS315-derived plasmid (*CEN LEU2*) containing *rpb1* deleted for the TL (*TL $\Delta$* ) and linearized at the deletion junction, allowing *in vivo* homologous recombination and resulting in a library of complete *rpb1* genes containing TL variants. The gap-repaired TL variants (Leu<sup>+</sup>) were

titered and plated at 200-300 colonies per plate to reduce inter-colony growth competition, and Leu<sup>+</sup> colonies were subsequently replica-plated to SC-LEU+5FOA (1mg/mL), and subsequently to additional selective and control media. Three independent biological replicate screens were performed. In each replicate, we pooled 6000 to 12000 colonies. Each cell pool was subjected to genomic DNA extraction and TL amplification by emulsion PCR. Amplification of the TL region was performed using Micellula DNA Emulsion & Purification (ePCR) Kit (Chimerx) per manufacturer's instructions. To minimize amplification bias, each sample was amplified in a 15-cycle ePCR reaction, purified and subject to another 13-15 cycle scale-up ePCR reactions. The two-step ePCR amplification protocol ensured sufficient yield of DNA for NGS sequencing while minimizing perturbation of the allele distribution in the DNA pool. The amplified samples were subject to Illumina HiSeq 2500 sequencing, and on average over 2 million reads were obtained from each replicate of a sample, with high reproducibility and minimal perturbation of the mutant distribution within the TL variant library (Supplementary Fig. 1d).

Allele frequency was subsequently measured by deep sequencing of the TL amplicons. To identify the mutations that were present for each set of paired-end reads, a codon-based alignment algorithm was developed to align each paired-end read set in which the overlapping substrings from both flanking regions agreed perfectly to the WT sequence. The purpose of our approach was to identify real variants using an expected set of mutant codons used in the programmed library synthesis from sequencing errors. A dynamic programming algorithm was applied so that an exact match of three letters was assigned a positive score, a mismatch of at least one letter in a codon was assigned a negative score, and the insertion or deletion of either one, two or three letters was assigned a constant negative score. The allele frequency was subsequently calculated from the mapped reads, and the phenotypic score of each TL variant was calculated by allele frequency change (normalized to WT) under each condition, as below:

$$f = \log \left[ \frac{f_{mut,sele}}{f_{mut,unsele}} \right] - \log \left[ \frac{f_{wt,sele}}{f_{wt,unsele}} \right]$$

Mutants with less than 200 reads in the transformed pool (SC-LEU) and allele frequency changes assessed from less than 50 reads from both conditions were excluded from further analyses. Median values from three independent biological replicates were used for fitness and phenotype scoring. Fitness score cutoff for lethality was estimated based on fitness scores (on SC-LEU and 5FOA) of 163 known viable TL and 16 known lethal mutants. Hierarchical clustering for generating phenotypic landscape was performed by Gene Cluster 3.0 using centered correlation<sup>73</sup>. Figures displaying structural information were generated using Pymol (<https://www.pymol.org/>).

### **Evolutionary analyses**

Eukaryotic RNA polymerase large subunit sequences were obtained from BLAST using *Sce* Rpb1 (Pol II), *Sce* Rpa190 (Pol I), and *Sce* Rpo31 (Pol III) sequences as queries. Sequences were assigned to Pol I, II, or III based on highest similarity when compared to each of the three query sequences, with prokaryotic sequences further filtered out. Multiple sequence alignments (MSAs) were generated by first applying CD-HIT<sup>74</sup> to cluster sequences so that the identity between sequences in different clusters was less than 90%, then applying MUSCLE<sup>75</sup> to obtain an alignment that contains one representative sequence from each cluster. The TL conservation score was generated

675 using Jalview<sup>76</sup> and plotted as a heatmap using Gene-E  
676 (<http://www.broadinstitute.org/cancer/software/GENE-E/index.html>).  
677

# References

1. 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**, 1876-82 (2001).
2. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-76 (2001).
3. Westover, K.D., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481-9 (2004).
4. 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**, 941-54 (2006).
5. Liu, X., Bushnell, D.A. & Kornberg, R.D. RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta* **1829**, 2-8 (2013).
6. Kaplan, C.D. Basic mechanisms of RNA polymerase II activity and alteration of gene expression in *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1829**, 39-54 (2013).
7. Zhang, J., Palangat, M. & Landick, R. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* **17**, 99-104 (2010).
8. Windgassen, T.A. et al. Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by *Escherichia coli* RNA polymerase. *Nucleic Acids Res* **42**, 12707-21 (2014).
9. Yuzenkova, Y. et al. Stepwise mechanism for transcription fidelity. *BMC Biol* **8**, 54 (2010).
10. 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 Research* **41**, 7048-7059 (2013).
11. 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**, 547-56 (2008).
12. Kireeva, M.L. et al. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell* **30**, 557-66 (2008).
13. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* **28**, 2663-76 (2014).
14. Malinen, A.M. et al. Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res* **40**, 7442-51 (2012).
15. 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**, 5863-70 (2014).
16. 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**, 882-93 (2013).
17. 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**, 767-75 (2013).



- 725 18. Vassylyev, D.G. et al. Structural basis for substrate loading in bacterial  
726 RNA polymerase. *Nature* **448**, 163-8 (2007).
- 727 19. Da, L.T., Wang, D. & Huang, X. Dynamics of pyrophosphate ion release  
728 and its coupled trigger loop motion from closed to open state in RNA  
729 polymerase II. *J Am Chem Soc* **134**, 2399-406 (2012).
- 730 20. Liu, B., Zuo, Y. & Steitz, T.A. Structures of E. coli sigmaS-transcription  
731 initiation complexes provide new insights into polymerase mechanism.  
732 *Proc Natl Acad Sci U S A* **113**, 4051-6 (2016).
- 733 21. Touloukhonov, I., Zhang, J., Palangat, M. & Landick, R. A central role of the  
734 RNA polymerase trigger loop in active-site rearrangement during  
735 transcriptional pausing. *Mol Cell* **27**, 406-19 (2007).
- 736 22. Larson, M.H. et al. Trigger loop dynamics mediate the balance between  
737 the transcriptional fidelity and speed of RNA polymerase II. *Proceedings of*  
738 *the National Academy of Sciences of the United States of America* **109**,  
739 6555-6560 (2012).
- 740 23. Seibold, S.A. et al. Conformational coupling, bridge helix dynamics and  
741 active site dehydration in catalysis by RNA polymerase. *Biochim Biophys*  
742 *Acta* **1799**, 575-87 (2010).
- 743 24. Silva, D.A. et al. Millisecond dynamics of RNA polymerase II translocation  
744 at atomic resolution. *Proc Natl Acad Sci U S A* **111**, 7665-70 (2014).
- 745 25. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S.A. Structural basis of  
746 transcriptional pausing in bacteria. *Cell* **152**, 431-41 (2013).
- 747 26. Wang, D. et al. Structural basis of transcription: backtracked RNA  
748 polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203-6 (2009).
- 749 27. Cheung, A.C. & Cramer, P. Structural basis of RNA polymerase II  
750 backtracking, arrest and reactivation. *Nature* **471**, 249-53 (2011).
- 751 28. Yuzenkova, Y. & Zenkin, N. Central role of the RNA polymerase trigger  
752 loop in intrinsic RNA hydrolysis. *Proc Natl Acad Sci U S A* **107**, 10878-83  
753 (2010).
- 754 29. Sosunova, E., Sosunov, V., Epshtein, V., Nikiforov, V. & Mustaev, A.  
755 Control of transcriptional fidelity by active center tuning as derived from  
756 RNA polymerase endonuclease reaction. *J Biol Chem* **288**, 6688-703  
757 (2013).
- 758 30. Miropolskaya, N. et al. Interplay between the trigger loop and the F loop  
759 during RNA polymerase catalysis. *Nucleic Acids Res* **42**, 544-52 (2014).
- 760 31. Esysunina, D. et al. Lineage-specific variations in the trigger loop modulate  
761 RNA proofreading by bacterial RNA polymerases. *Nucleic Acids Res* **44**,  
762 1298-308 (2016).
- 763 32. Kettenberger, H., Armache, K.J. & Cramer, P. Complete RNA polymerase  
764 II elongation complex structure and its interactions with NTP and TFIS.  
765 *Mol Cell* **16**, 955-65 (2004).
- 766 33. Lennon, C.W. et al. Direct interactions between the coiled-coil tip of DksA  
767 and the trigger loop of RNA polymerase mediate transcriptional regulation.  
768 *Genes Dev* **26**, 2634-46 (2012).

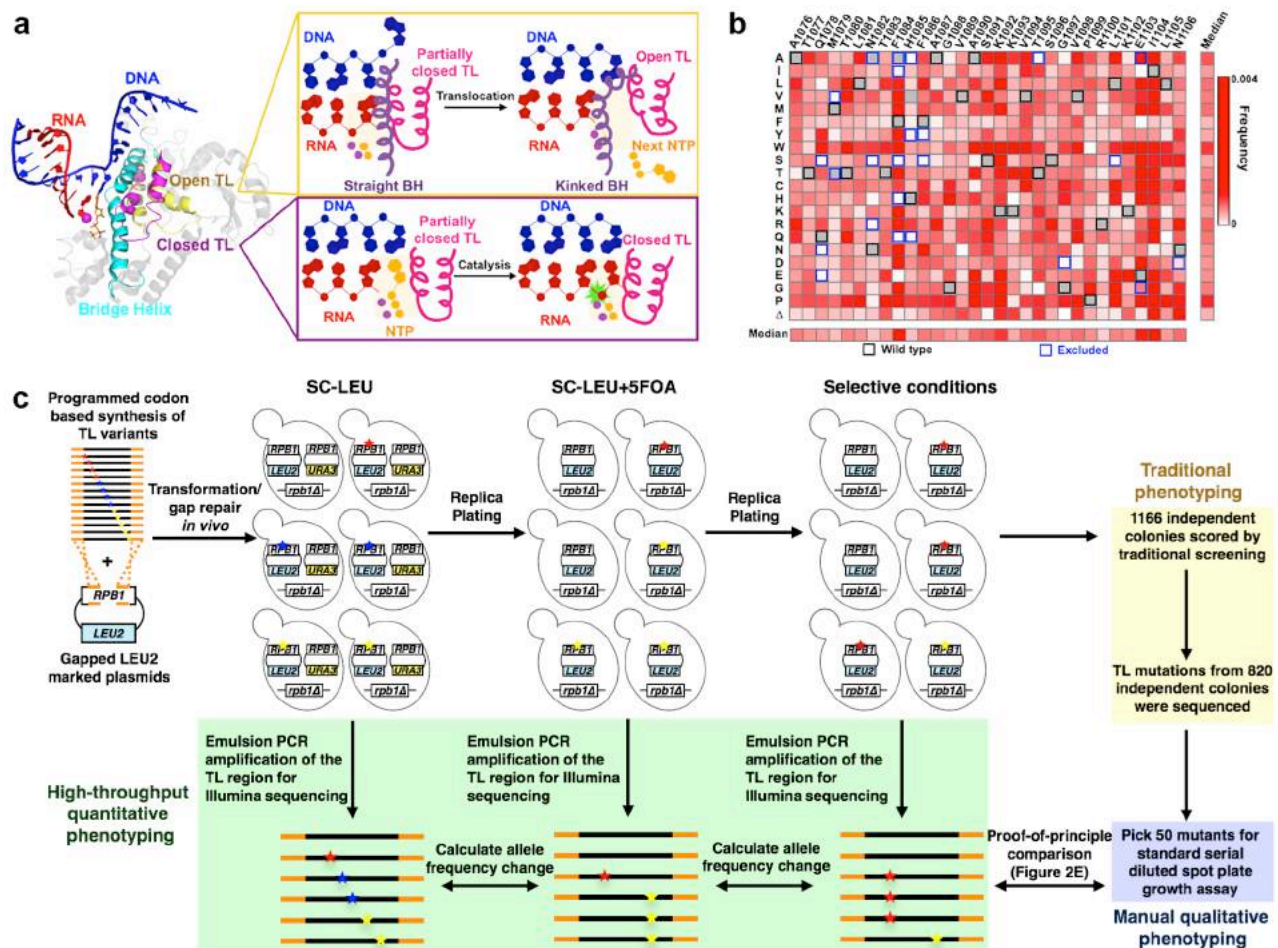
- 769 34. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The  
770 ratcheted and ratchetable structural states of RNA polymerase underlie  
771 multiple transcriptional functions. *Mol Cell* **57**, 408-21 (2015).
- 772 35. Hein, P.P. et al. RNA polymerase pausing and nascent-RNA structure  
773 formation are linked through clamp-domain movement. *Nature Structural*  
774 *& Molecular Biology* **21**, 794-802 (2014).
- 775 36. Kaplan, C.D., Jin, H., Zhang, I.L. & Belyanin, A. Dissection of Pol II trigger  
776 loop function and Pol II activity-dependent control of start site selection in  
777 vivo. *PLoS Genet* **8**, e1002627 (2012).
- 778 37. Malagon, F. et al. Mutations in the *Saccharomyces cerevisiae* RPB1 gene  
779 conferring hypersensitivity to 6-azauracil. *Genetics* **172**, 2201-9 (2006).
- 780 38. Irvin, J.D. et al. A genetic assay for transcription errors reveals multilayer  
781 control of RNA polymerase II fidelity. *PLoS Genet* **10**, e1004532 (2014).
- 782 39. Dangkulwanich, M. et al. Complete dissection of transcription elongation  
783 reveals slow translocation of RNA polymerase II in a linear ratchet  
784 mechanism. *Elife* **2**, e00971 (2013).
- 785 40. Braberg, H. et al. From structure to systems: high-resolution, quantitative  
786 genetic analysis of RNA polymerase II. *Cell* **154**, 775-88 (2013).
- 787 41. Viktorovskaya, O.V. et al. Divergent contributions of conserved active site  
788 residues to transcription by eukaryotic RNA polymerases I and II. *Cell Rep*  
789 **4**, 974-84 (2013).
- 790 42. Mejia, Y.X., Nudler, E. & Bustamante, C. Trigger loop folding determines  
791 transcription rate of *Escherichia coli*'s RNA polymerase. *Proc Natl Acad*  
792 *Sci U S A* **112**, 743-8 (2015).
- 793 43. Fowler, D.M. & Fields, S. Deep mutational scanning: a new style of protein  
794 science. *Nat Methods* **11**, 801-7 (2014).
- 795 44. McLaughlin, R.N., Jr., Poelwijk, F.J., Raman, A., Gosal, W.S. &  
796 Ranganathan, R. The spatial architecture of protein function and  
797 adaptation. *Nature* **491**, 138-42 (2012).
- 798 45. Araya, C.L. et al. A fundamental protein property, thermodynamic stability,  
799 revealed solely from large-scale measurements of protein function. *Proc*  
800 *Natl Acad Sci U S A* **109**, 16858-63 (2012).
- 801 46. Kim, I., Miller, C.R., Young, D.L. & Fields, S. High-throughput analysis of  
802 in vivo protein stability. *Mol Cell Proteomics* **12**, 3370-8 (2013).
- 803 47. Zhai, W. et al. Synthetic antibodies designed on natural sequence  
804 landscapes. *J Mol Biol* **412**, 55-71 (2011).
- 805 48. Van den Brulle, J. et al. A novel solid phase technology for high-  
806 throughput gene synthesis. *Biotechniques* **45**, 340-3 (2008).
- 807 49. Tewhey, R. et al. Microdroplet-based PCR enrichment for large-scale  
808 targeted sequencing. *Nat Biotechnol* **27**, 1025-31 (2009).
- 809 50. Williams, R. et al. Amplification of complex gene libraries by emulsion  
810 PCR. *Nat Methods* **3**, 545-50 (2006).
- 811 51. Barnes, C.O. et al. Crystal Structure of a Transcribing RNA Polymerase II  
812 Complex Reveals a Complete Transcription Bubble. *Mol Cell* **59**, 258-69  
813 (2015).

- 814 52. Cabart, P., Jin, H., Li, L. & Kaplan, C.D. Activation and reactivation of the  
815 RNA polymerase II trigger loop for intrinsic RNA cleavage and catalysis.  
816 *Transcription* **5**, e28869 (2014).
- 817 53. Niyogi, S.K., Feldman, R.P. & Hoffman, D.J. Selective effects of metal  
818 ions on RNA synthesis rates. *Toxicology* **22**, 9-21 (1981).
- 819 54. Niyogi, S.K. & Feldman, R.P. Effect of several metal ions on  
820 misincorporation during transcription. *Nucleic Acids Res* **9**, 2615-27  
821 (1981).
- 822 55. Dyson, H.J., Wright, P.E. & Scheraga, H.A. The role of hydrophobic  
823 interactions in initiation and propagation of protein folding. *Proc Natl Acad*  
824 *Sci U S A* **103**, 13057-61 (2006).
- 825 56. Cheung, A.C., Sainsbury, S. & Cramer, P. Structural basis of initial RNA  
826 polymerase II transcription. *EMBO J* **30**, 4755-63 (2011).
- 827 57. Rudolph, H.K. et al. The yeast secretory pathway is perturbed by  
828 mutations in PMR1, a member of a Ca<sup>2+</sup> ATPase family. *Cell* **58**, 133-45  
829 (1989).
- 830 58. Mandal, D., Woolf, T.B. & Rao, R. Manganese selectivity of pmr1, the  
831 yeast secretory pathway ion pump, is defined by residue gln783 in  
832 transmembrane segment 6. Residue Asp778 is essential for cation  
833 transport. *J Biol Chem* **275**, 23933-8 (2000).
- 834 59. Kaster, B.C., Knippa, K.C., Kaplan, C.D. & Peterson, D.O. RNA  
835 Polymerase II Trigger Loop Mobility: Indirect Effects of Rpb9. *J Biol Chem*  
836 (2016).
- 837 60. Vassylyev, D.G. et al. Crystal structure of a bacterial RNA polymerase  
838 holoenzyme at 2.6 Å resolution. *Nature* **417**, 712-9 (2002).
- 839 61. Hirata, A., Klein, B.J. & Murakami, K.S. The X-ray crystal structure of RNA  
840 polymerase from Archaea. *Nature* **451**, 851-4 (2008).
- 841 62. Tan, L., Wiesler, S., Trzaska, D., Carney, H.C. & Weinzierl, R.O. Bridge  
842 helix and trigger loop perturbations generate superactive RNA  
843 polymerases. *J Biol* **7**, 40 (2008).
- 844 63. Weinzierl, R.O. The nucleotide addition cycle of RNA polymerase is  
845 controlled by two molecular hinges in the Bridge Helix domain. *BMC Biol*  
846 **8**, 134 (2010).
- 847 64. Da, L.T. et al. Bridge helix bending promotes RNA polymerase II  
848 backtracking through a critical and conserved threonine residue. *Nat*  
849 *Commun* **7**, 11244 (2016).
- 850 65. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. & Cramer, P. RNA  
851 polymerase I structure and transcription regulation. *Nature* **502**, 650-5  
852 (2013).
- 853 66. Huang, X. et al. RNA polymerase II trigger loop residues stabilize and  
854 position the incoming nucleotide triphosphate in transcription. *Proc Natl*  
855 *Acad Sci U S A* **107**, 15745-50 (2010).
- 856 67. Castro, C. et al. Nucleic acid polymerases use a general acid for  
857 nucleotidyl transfer. *Nat Struct Mol Biol* **16**, 212-8 (2009).

- 858 68. Jin, H. & Kaplan, C.D. Relationships of RNA polymerase II genetic  
859 interactors to transcription start site usage defects and growth in  
860 *Saccharomyces cerevisiae*. *G3 (Bethesda)* **5**, 21-33 (2014).
- 861 69. Winston, F., Dollard, C. & Ricupero-Hovasse, S.L. Construction of a set of  
862 convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C.  
863 *Yeast* **11**, 53-5 (1995).
- 864 70. Puig, O. et al. The tandem affinity purification (TAP) method: a general  
865 procedure of protein complex purification. *Methods* **24**, 218-29 (2001).
- 866 71. Schmitt, M.E., Brown, T.A. & Trumpower, B.L. A rapid and simple method  
867 for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids*  
868 *Res* **18**, 3091-2 (1990).
- 869 72. Ranish, J.A. & Hahn, S. The yeast general transcription factor TFIIA is  
870 composed of two polypeptide subunits. *J Biol Chem* **266**, 19320-7 (1991).
- 871 73. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering  
872 software. *Bioinformatics* **20**, 1453-4 (2004).
- 873 74. Li, W., Jaroszewski, L. & Godzik, A. Tolerating some redundancy  
874 significantly speeds up clustering of large protein databases.  
875 *Bioinformatics* **18**, 77-82 (2002).
- 876 75. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy  
877 and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).
- 878 76. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J.  
879 Jalview Version 2--a multiple sequence alignment editor and analysis  
880 workbench. *Bioinformatics* **25**, 1189-91 (2009).

881  
882  
883





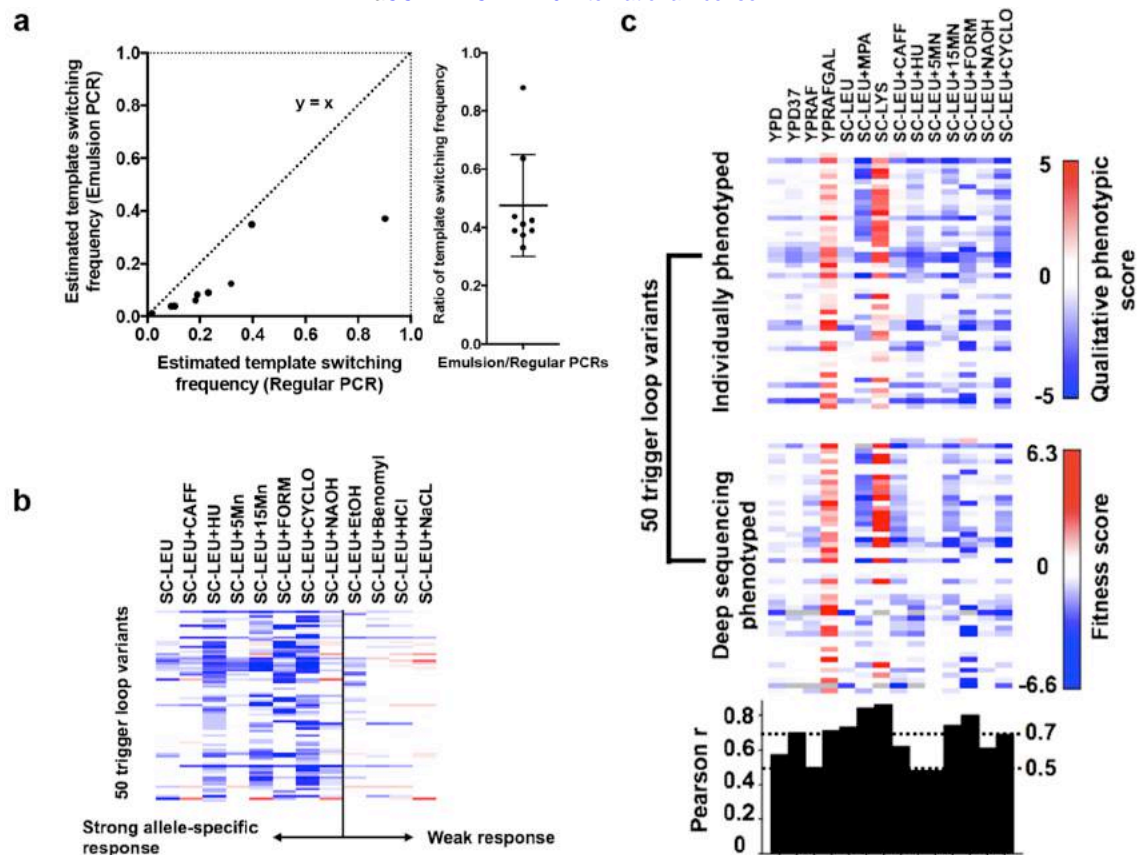
**Figure 1: Establishment of a high-throughput platform for phenotyping comprehensive TL single variant library**

Figure 1a: Multiple TL functions are underpinned by its mobile nature. Structures of open (PDB:5C4J) and closed TLs (PDB:2E2H) are shown in the context of surrounding domains. DNA (blue), RNA (red), Bridge Helix (cyan), Closed TL (magenta) and Open TL (yellow) are shown in cartoon rendering. The open TL has been proposed to allow Pol II translocation while the closed TL has been shown to facilitate catalysis (right panel).

Figure 1b: Mutational coverage of the TL variant library is shown as heatmap illustrating the allelic frequencies of single substitution variants (WT amino acids and positions labeled on x axis; amino acid substitutions on y axis). The WT amino acids for each position are noted with black boxes, and mutants excluded from library synthesis are noted using blue boxes.

Figure 1c: Schematic representation of experimental approach. Stars of different colors represent distinct substitutions. The TL variant library PCR amplicon (encoding Rpb1 amino acids 1076-1106) flanked by *RPB1* TL flanking sequence (orange) was co-transformed with a linearized *LEU2* *CEN* plasmid containing an *rpb1* gene with the TL deleted, allowing construction of full-length *RPB1* (with TL variants) by *in vivo* homologous recombination. Heterozygous *Leu*<sup>+</sup> transformants were replica-plated onto SC-LEU+5FOA to select against the WT *RPB1* (*URA3* *CEN*) plasmid and to create TL variant pools. TL variant pools were subsequently replica-plated to different selective conditions for either traditional individual colony screening or high-throughput phenotyping using deep sequencing. For the latter, replica-plated colonies were pooled.

908 for genomic DNA extraction, and the TL region was amplified by emulsion PCR to  
909 prepare templates for deep sequencing.  
910  
911  
912



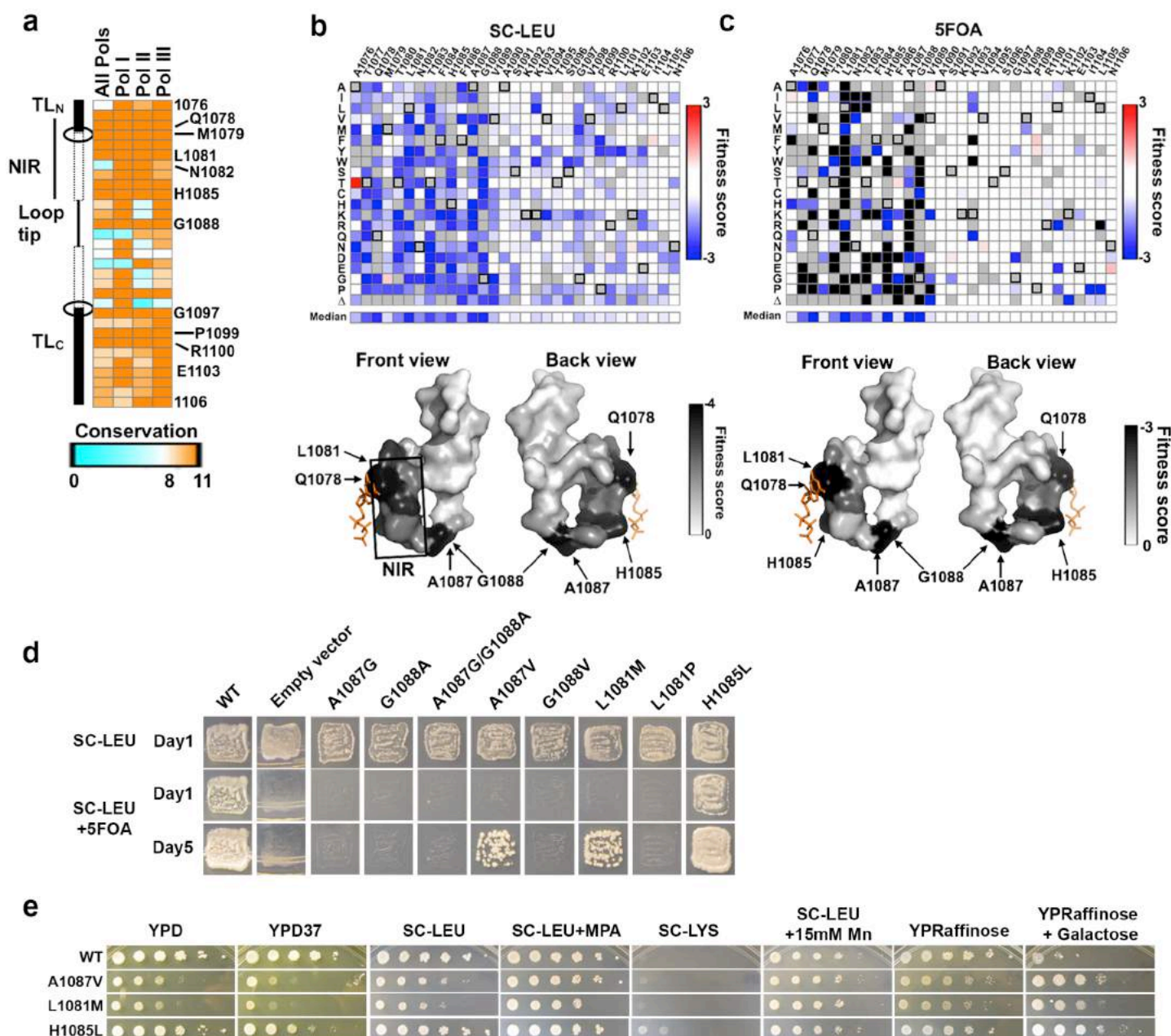
## Figure 2: Quality controls for the high-throughput phenotyping approach

Figure 2a: Comparison of estimated template switching frequency in regular and emulsion PCR conditions. Template switching was estimated by the ratio  $(\text{Freq}^{\text{Double}}) / (\text{Freq}^{\text{Single1}} \times \text{Freq}^{\text{Single2}})$  for all the possible double mutants combined from five spiked-in single mutants.

Figure 2b: Additional growth conditions were employed to increase resolution for distinguishing similar TL alleles. Growth scores for 50 individually isolated TL mutants (y axis) under 12 growth conditions (x axis), as determined by standard serial dilution plate phenotyping, are shown as a heatmap (Supplementary figure 2). Positive values shown in red indicate better growth than WT and negative values in blue indicate worse growth than WT.

Figure 2c: High-throughput quantitative phenotyping results are consistent with individual phenotyping of variants. Top heatmap shows qualitative growth scores (as in Fig. 2b) of 50 individually phenotyped TL variants on y axis (supplementary figure 2) with selective conditions on the x axis. Deep sequencing results for the same mutants using median of fitness defects from three independent high-throughput screens are shown in the middle panel. Pearson r calculated to show the correlation between each condition from the two datasets were shown in the bottom panel.





**Figure 3: The TL fitness landscape distinguishes highly conserved TL residues and reveals high mutational sensitivity in the nucleotide interacting region (NIR) and the Alanine-Glycine linker.**

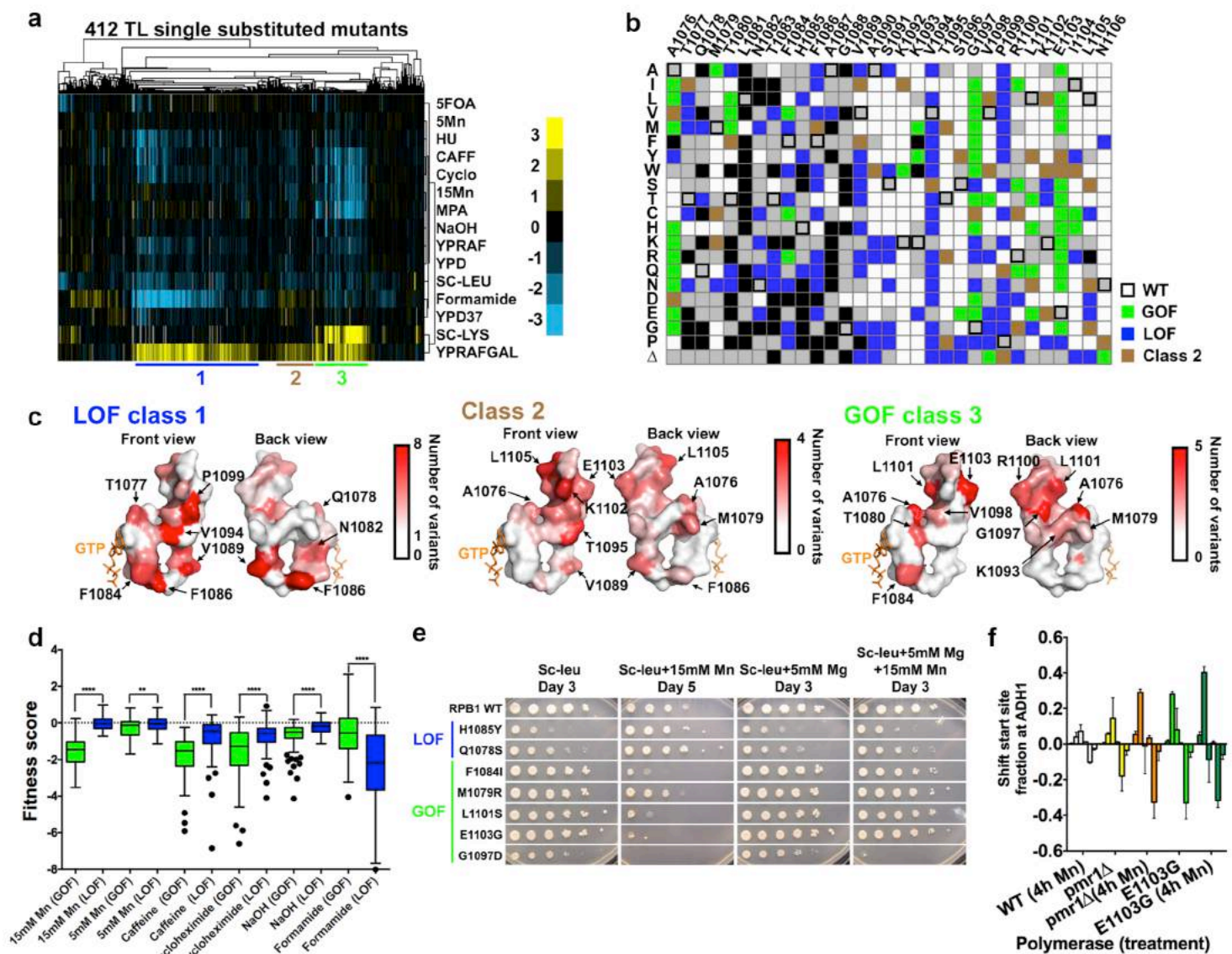
Figure 3a: Conservation heatmap of TL residues in eukaryotic RNA polymerases. The conservation scores were extracted from a multiple sequence alignment, including 182 Pol II, 59 Pol I, and 111 Pol III sequences.

Figure 3b: Fitness defects of TL variants in the heterozygous state are shown as a heatmap. Unavailable data points are in grey squares. WT residues at indicated positions are in black box. Surface representation (bottom panel) of the TL structure (PDB:2E2H) is shaded by the median fitness value for all available variants at each position, in a gradient of white (rare defects) to black (common defects). The position of matched GTP substrate is shown in orange stick representation.

Figure 3c: General fitness defects of TL variants upon removal of WT *RPB1*. Fitness defects predicted to result in lethality shown in black. Surface representation (bottom



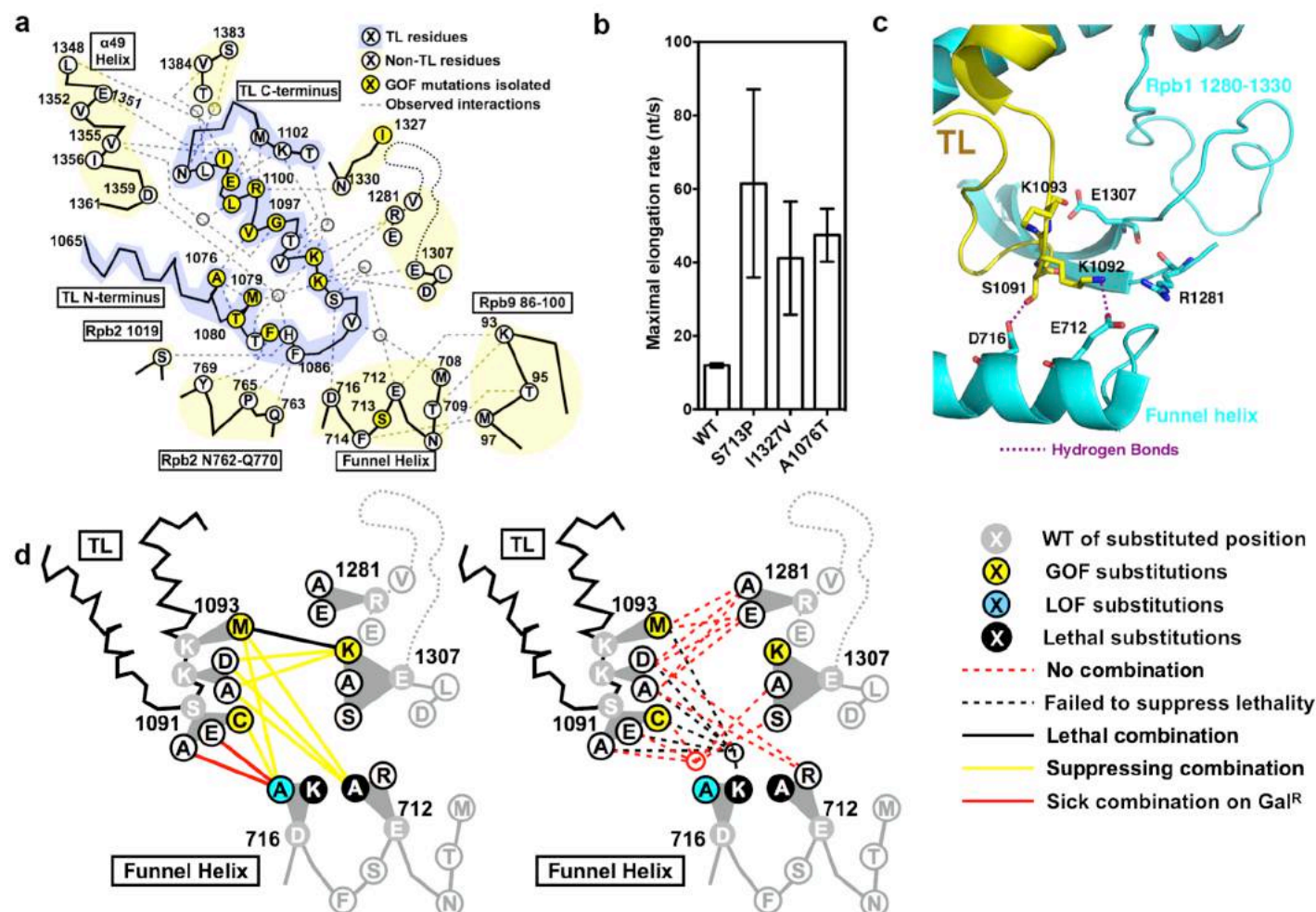
panel) of the TL structure is shaded by the median fitness value of all available variants at each position, in a gradient of white (rare defects) to black (defects common).  
 Figure 3d: Complementation abilities of variants in the difficult-to-substitute TL positions (L1081, A1087, G1088) or unexpected TL variants (H1085L) assayed by plasmid shuffling of individual strains. Ability to grow on SC-LEU+5FOA indicates complementation of essential functions of *RPB1*. SC-LEU medium is the control state where WT *RPB1* is present.  
 Figure 3e: Transcription-linked phenotypes of the viable mutants from the difficult-to-substituted residues L1081M, A1087V or unexpected TL variant H1085L.



**Figure 4: Three distinct TL mutant classes, revealed from TL phenotypic landscape, have specific distribution on the TL structure and distinct stress response profiles**

Figure 4a: Hierarchical clustering of 412 single TL variants' (y axis) phenotypes (calculated as in Figure 1c) under 14 different conditions (x axis) reveals distinct mutant classes. Positive (yellow) and negative (blue) fitness scores are shown as a heatmap. Mutant classes (clusters) are annotated by colored lines beneath the heatmap. Figure 4b: Distribution of three major mutant classes is shown in a single substitution variant heatmap. Class 1 (GOF) mutants are shown in green; Class 2 mutants are shown in brown and Class 3 (LOF) mutants are shown in blue. Figure 4c: Distribution of different mutant classes on the TL structure. TL is shown in surface and colored in the gradient from white to red by the number of clustered mutants at each position. Figure 4d: Differential stress responses in GOF and LOF mutants. GOF mutants are more sensitive to Mn, caffeine and cycloheximide, whereas LOF mutants are more sensitive to formamide. \*\*p<0.01, \*\*\*\*p<0.0001 (Two-tailed unpaired t-test). Figure 4e: Differential Mn<sup>2+</sup> sensitivity and its suppression by Mg<sup>2+</sup> for selected TL variants representative of mutant classes.

978 Figure 4f:  $Mn^{2+}$  effects on different mutants' transcription start sites (TSSs) distribution at  
979 ADH1, determined by primer extension analysis. TSSs at ADH1 are distributed in a  
980 range of positions and were divided into six bins for quantitation: from upstream (left) to  
981 downstream (right). Change of TSSs (normalized to untreated WT) is calculated by the  
982 fractions. Average and standard deviation of three experimental replicates are shown  
983 with plotted by bar graph with error bars.  
984



# **Figure 5: Functional contribution of TL tip and funnel helix to proper TL dynamics**

Figure 5a: Observed and predicted interactions between TL and TL-proximal domains.

TL schematic is shown with residues identified by single-letter amino acid code and positions of interest with positions of isolated GOF mutants color coded in yellow.

Observed TL interactions with other Rpb1 domains from structures or simulation studies

are shown in the grey dashed lines.

Figure 5b: Maximal elongation rates of Pol II WT and genetically GOF mutants S713P,

I1327V and A1076T.

Figure 5c: Observed interactions between open TL tip and TL adjacent charged residues

(PDB: 5C4X).

Figure 5d: Genetic interactions between TL tip and proximal Rpb1 domains. Schematic

of TL and the adjacent domains are shown in lines, with positions of interests shown in

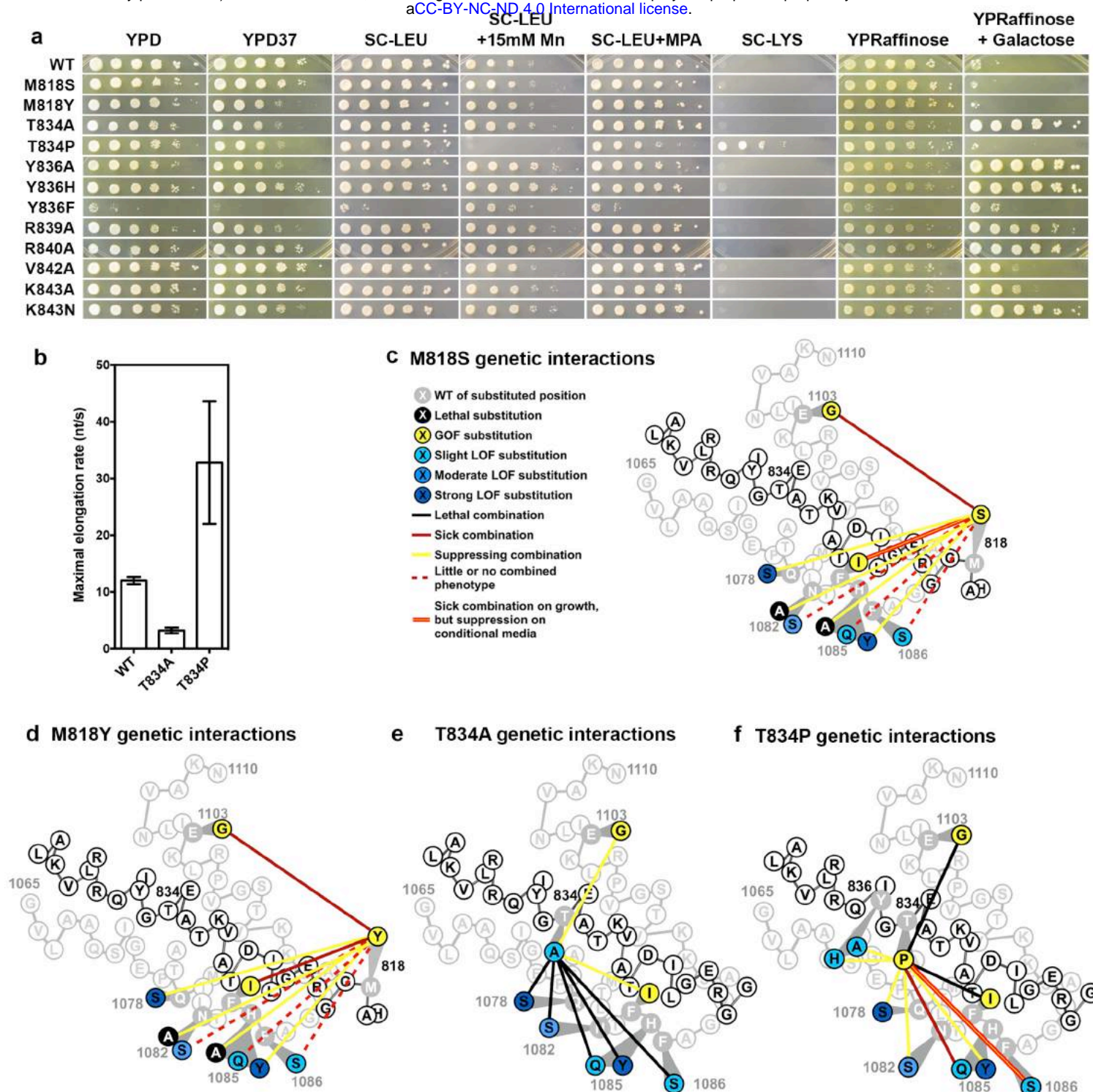
single-letter code. Substituted residues are shown in grey, with substituting amino acids

shown in white, blue or yellow filled circles based on single mutant phenotypes

(Supplementary Figure 3). Double substitution phenotypes are shown as colored lines

connecting the two relevant single substitutions.





**Figure 6: Functional interplay between TL and Bridge Helix (BH)**

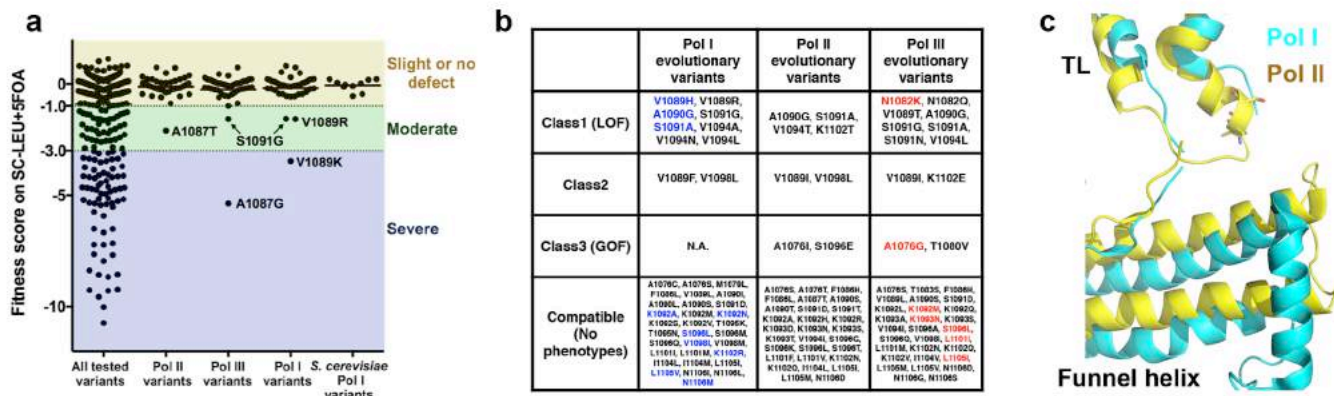
Figure 6a: Transcription-linked phenotypes of BH single-substituted mutants. M818S and M818Y are substitutions in a predicted BH N-terminal hinge; others are substitutions in predicted BH C-terminal hinge positions or additional C-terminal substitutions.

Figure 6b: Maximal elongation rate of BH variants T834A and T834P *in vitro*.

Figure 6c: Genetic interactions between BH M818S and TL substitutions. M818S suppressed (yellow lines) the strong LOF TL variants (dark blue) but not the slight and moderate LOF TL variants (light blue), and M818S was synthetic sick (red line) with the GOF TL variants (yellow).



Figure 6d: Genetic interactions between BH M818Y and TL substitutions. Similar to M818S genetic interaction with TL variants (Figure 6c), M818Y suppressed (yellow lines) the strong LOF TL variants (dark blue) but not the slight and moderate LOF TL variants (light blue), and M818Y was synthetic sick (red line) with GOF TL variants (yellow). Figure 6e: Genetic interactions between BH T834A and TL substitutions. T834A suppressed (yellow lines) the GOF TL variants and was synthetic lethal with all the tested LOF TL variants (blue). Figure 6f: Genetic interactions between BH T834P and TL or BH. Similar to M818 variants (Figure 6c, 6d), T834P suppressed the strong and moderate LOF TL variants (dark blue) but was synthetic sick with the slight LOF TL variants (light blue), while synthetic lethal with the GOF TL variants (yellow). T834P was also suppressed (yellow) by two LOF BH mutants Y836A/H.



**Figure 7: Phenotypic analyses of evolutionary variants reveal context dependent function of many TL residues**

Figure 7a: General growth fitness defects of the TL single-substituted variants observed in TL across Pol I, II, III evolution including 38 Pol II, 42 Pol I and 42 Pol III amino acid variants relative to Sce Pol II.

Figure 7b: Evolutionary TL variants in three mutant classes from the TL phenotypic landscape (Figure 4a, 4b). Existing variants from Sce Pol I are colored in blue, and existing variants from Sce Pol III are colored in red. Sce Pol I has three substitutions (V1089H, A1090G and S1091A) that cause LOF in the Pol II context; Sce Pol III has one substitution (A1076G) classified as GOF and one substitution (N1082K) classified as LOF.

Figure 7c: Difference in the funnel helix positioning (relative to TL) in Pol I and Pol II. Cartoon representation of TL/funnel helix from Pol I and Pol II are shown in cyan and yellow, respectively (PDB: 5C4J and 2VUM).

1044 **Table 1 Plate phenotypes employed for the screening Pol II alleles *in vivo***

Phenotype	Affected Gene/Reporter Allele; Pol II mutant class affected	WT growth	Mutant growth
Sensitivity to 5FOA	Detects ability of <i>rpb1 LEU2</i> plasmid to complement <i>rpb1Δ</i>	Resistant to drug. <i>RPB1 LEU2</i> plasmid complements <i>rpb1Δ</i>	Sensitivity to drug (Partial or no complementation of <i>rpb1Δ</i> by <i>rpb1 LEU2</i> )
Suppressor of Ty (Spt <sup>-</sup> )	<i>lys2-128Δ</i> ; reports on chromatin defects and start site selection. Specific class of GOF Pol II mutants.	Lysine auxotroph (Lys <sup>-</sup> )	Lysine prototroph (Lys <sup>+</sup> )
Mycophenolic acid sensitivity (MPA <sup>S</sup> )	<i>IMD2</i> expression required for resistance; reports on start site selection. Specific classes of GOF and LOF Pol II mutants.	Resistant to drug	Sensitivity to drug for GOF mutants, relative resistance for LOF mutants.
Modulation of transcriptional readthrough at <i>gal10Δ56</i> (Gal <sup>R</sup> )	<i>gal10Δ56</i> ; likely reports on termination, mRNA processing and initiation. It was found widely in LOF Pol II mutants, some GOF.	Moderate sensitivity to galactose (Gal <sup>S</sup> )	Resistance to galactose (Gal <sup>R</sup> )

1045