Function and evolution of B-Raf loop dynamics relevant to cancer recurrence under drug inhibition

Gregory A. Babbitt[1]*, Miranda L. Lynch[2], Matthew McCoy[3], Ernest P. Fokoue[4], and André O. Hudson[1]

[1]Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester NY, USA 14623
[2]Hauptmann-Woodward Medical Research Institute, Buffalo NY, USA 14203
[3]Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington DC, USA 20007
[4]School of Mathematical Sciences, Rochester Institute of Technology, Rochester NY, USA 14623

Contact Information –Gregory A. Babbitt (gabsbi@rit.edu)

Short title – cancer recurrence and protein dynamics

Authors' contributions – GAB and EPF designed the statistical method and software. GAB, MLL, MM, designed the study and all author's interpreted the data and wrote the manuscript.

Abstract

Oncogenic mutations in the kinase domain of the B-Raf protein have long been associated with cancers involving the RAF-MEK-ERK pathway. One constitutive ERK activating mutation in B-Raf, the V600E (valine to glutamate) replacement occurring adjacent to a site of threonine phosphorylation (T599) occurs in many types of cancer, and in a large percentage of certain cancers, such as melanoma. Because natural ATP binding activity and the V600E mutation are both known to alter the physical behavior of the activation loop in the ATP binding domain, this system is especially amenable to comparative functional analyses using molecular dynamics of drug classes and genetic variants at all-atom resolution. Here, we employ machine learning enabled identification of functionally conserved protein dynamics to compare how disease pertinent molecular variations impact conserved loop dynamics during the binding of four ATP competitive B-Raf inhibitors (sorafenib, vemurafenib, dabrafenib, and PLX7904) that differ in their effectiveness in preventing cancer recurrence. We demonstrate that drug development targeting B-Raf has progressively moved towards ATP competitive inhibitors that demonstrate less tendency to mimic the functional dynamic shifts associated with ATP binding in this domain. We argue this functional dynamic mimicry in first generation B-Raf inhibitors increases the side effect of hyperactivation (i.e. inducing MAPK activation in non-tumorous cells in the absence of secondary mutation). Within the context of the binding interaction of each inhibitor, we compare the functional dynamic impacts of V600E and other sensitizing and drug resistance causing mutations in the activation and P-loops, confirming sites of low mutational tolerance in these two functional regions. Lastly, we investigate V600E sensitivity of B-Raf loop dynamics in an evolutionary context, demonstrating that while it probably shares partial origin with its early evolution in primitive eukaryotes, the functional sensitivity to V600E was secondarily increased during early jawed vertebrate evolution.

Author Summary

Our lab has recently developed a combined simulation-based and machine learning method and user-friendly software for identifying dynamic motions in proteins that are preserved over evolutionary time scales (i.e. are functionally conserved). Our method also can assess how genetic or chemical variation affects these functionally conserved motions. Here, we utilize this approach to computationally demonstrate how one of the most common mutations that arise in tumors, the V600E B-Raf protein mutation, disrupts the functioning of protein loop dynamics responsible for normal regulation of the oncogenic B-Raf protein and its growth signaling pathway. We demonstrate that cancer recurrence in the absence of new mutations may be caused by drug interactions impacting these dynamics. We propose that mimicry of the functional dynamics of natural chemical activators (i.e. agonists) by many cancer targeting drugs may be a potent causal factor in recurrence of cancer in the absence of secondary mutation. By combining comparative modeling of molecular dynamics and analyses of molecular evolution, we also demonstrate that the unusual sensitivity of B-Raf protein regulatory loop dynamics to this mutation has an ancient evolutionary origin.

Keywords – oncogene, mutation, B-Raf, inhibitor, molecular dynamics, molecular evolution, machine learning, dynamic mimicry

Introduction

From both a patient's and doctor's perspective, perhaps the most dreaded aspect of many cancers are their ability to undergo recurrence following therapeutic intervention. The recurrence of cancer can have obvious negative impact on patient psychology [1,2], doctor-patient relations [3] and questions the very concept of a 'cure' for cancer. The molecular mechanisms and somatic evolutionary processes that underlie the phenomena of recurrence are complex and do not act singularly within any given tumor, thereby greatly exacerbating the clinical difficulty of treatment and post-treatment monitoring. There are currently three molecular mechanistic frameworks for potentially explaining cancer recurrence. The traditional view of 'mutational deregulation' invokes a progressive molecular evolutionary Fisherian or 'runaway' process whereby tumor progression begins with one or several 'driver mutations' that start a cascade of secondary mutation and subsequent deregulation of growth and proliferation followed by eventual chromosomal disruption through aneuploidy and chromothrypsis [4]. Under this conceptual framework, cancer progression is largely due to a mutational process that progressively deregulates cell proliferation and frames cancer recurrence as a problem analogous to a 'whack a mole' game that gets more difficult as the game progresses. A second molecular evolutionary perspective that has gained recent popularity invokes Darwinian natural selection at the somatic level of the tumor microenvironment [5,6]. Under this view, the cancer treatment regime, in creating a local tumor microenvironment that is selectively detrimental to the survival and proliferation of cancer cells, is thought to create a selective microenvironment that favors more robust and treatment-resistant cancers that can recur if therapy is not modulated accordingly. Cancer recurrence under this new paradigm is framed as a game of strategy like chess, where the competitive environment changes fluidly with the progression of the game. Both of these conceptual frameworks for our understanding of cancer recurrence rely upon disruption of mutation-selection balance at a tissue level [7,8] acting through the complementary mechanisms of somatic mutation and apoptosis [9].

A third and less common cause of cancer recurrence, known as the 'hyperactivation paradox' does not rely upon disruption of selection-mutation balance and can be invoked to explain cancer recurrence in the absence of secondary mutations. Hyperactivation is observed when some cancer inhibitors have opposing effects in tumor and normal wildtype cells, repressing growth pathways in tumor tissues while paradoxically activating them in normal healthy tissue [10–12]. The mechanism of hyperactivation may be less of a paradox when one considers that small molecule inhibitors are often targeting the binding sites of natural agonists (i.e. signal pathway activators) in both tumor and wildtype cells. In growth activated tumor cells, these drugs will competitively bind the target protein, thus blocking the functional role of the natural agonist in the pathway. However, in growth deactivated cells, these drugs might mimic certain functional aspects of the agonist, allowing partial activation of growth pathways in the presence of the drugs. While not a 'game' that most would choose to play, this effect of functional mimicry in some drugs might be analogous to 'poking a sleeping bear', where cancer recurrence can be caused in the absence of secondary mutation by a drug's unintended signal activation of the growth pathway in nearby normal cells.

Proteins involved in growth signaling pathways, such as mitogen-activeted protein kinases (MAPK), are often very labile in structure, involving highly dynamic regions of intrinsic disorder (e.g. activation loops) that facilitate complex dynamic shifts between functional states that ultimately govern interactions with other proteins on the pathway [13,14]. Mutations that potentially alter function in

114 highly dynamic loop regions are often implicated in many cancers. A well-studied oncogenic system
115 exists in the ATP binding domain of the B-Raf protein, a serine/threonine-protein kinase that activates
116 the MAPK (i.e. RAF-MEK-ERK) growth pathway (Figure 1). B-Raf is a member of the Raf family of
117 serine/threonine kinases, along with A-Raf and C-Raf, and is the member with the highest basal activity
118 of the three [15]. B-Raf is a 766 amino acid protein with gene on chromosome 7 (7q34). There are 3
119 primary conserved domains for B-Raf, CR1, 2, and 3. CR3 (residues 443-735) is the domain that contains
120 the enzymatic ATP kinase domain (KD, residues 457-714). B-Raf regulation functions primarily by the
121 changes in the position and dynamics in several loop regions upon interaction of ATP with the
122 hydrophobic binding pocket formed by V471, A481, L514, T529, W531, and C532. Within the KD, the
123 glycine-rich P-loop (ATP-phosphate stabilizing loop) spans residues 464-471 and functions to anchor ATP
124 during enzymatic activity. The activation segment or activation loop (residues 594-623) within the KD is
125 partially disordered and contains the serine and threonine phosphorylation sites (S602 and T599). Thus,
126 the loop regions primarily affected by ATP binding interaction are the ATP phosphate stabilizing P-loop
127 at residues 464-471, the catalytic loop at residues 574-581 that captures and transfers the ATP γ-
128 phosphate, and the activation loop segment and DFG motif at residues 594-623 which breaks
129 hydrophobic interactions with the P-loop (i.e. activating the KD) with negative charge during
130 phosphorylation of T599. Activating mutations in the KD can often lead to disrupted conformation and
131 constitutive ERK activation and many mutations in this region are associated with loss of regulatory
132 function and/or abnormal dimerization [16]. In particular, a V600E mutation in the activation loop
133 causing the constitutive activation of the B-Raf signaling is one of the most common mutations in many
134 type of cancer, as well as a high percentage of certain cancers like melanoma. B-Raf mutants appear in
135 multiple cancers, notably in melanoma, and colorectal, but also less frequently in non-small cell lung
136 cancer, breast, ovarian, and several others [14,17,18]. Greater than 50% melanomas harbor B-Raf
137 activating mutations, and of these 90% involve the three bases of codon 600, and of these 90% are the
138 single nucleotide missense mutation at position 1799 resulting in V600E substitution (although V600 K,
139 D, R have also been less commonly observed [19]).

140 ***Fig 1. Overview of role of BRAF in mitogen-activated protein kinase pathway (MAPK) pathway.***
141 *Growth factor/s binds to the receptor (EGFR) on the cell surface which leads to phosphorylation and*
142 *subsequent activation. The scaffold proteins (GRB2 and SOS) facilitate the removal of GDP from RAS. RAS*
143 *then binds GTP which leads to activation. Activated RAS undergoes a conformational change which leads*
144 *to the phosphorylation and activation of BRAF, MEK and ERK. Activated ERK is shuttled to the nucleus*
145 *where it is involved in recruiting transcription factors involved in the up regulation of genes involved in*
146 *cell proliferation, cell survival, cell motility and cell invasion. The BRAF V600E mutation allows for*
147 *constitutive activation of BRAF to facilitate downstream signaling through MEK and ERK circumventing*
148 *upstream activation/regulation.*

149 Several first generation B-Raf inhibitors designed to target V600E mutants cells have been
150 recently discovered to induce hyperactivation in normal wildtype cells in the absence of V600E [10–12].
151 Now, a new generation of B-Raf inhibitors that break the hyperactivation paradox are currently under
152 development and clinical trial [20]. Raf inhibitors are generally designed to occupy the ATP binding
153 pocket preventing phosphorylation. Why this region suffers from such low mutational tolerance, and
154 why some Raf inhibitors prove vulnerable to hyperactivation while others do not, are currently open
155 questions. A comparative analysis of the all-atom molecular dynamics (MD) of the B-Raf KD in both its
156 wildtype and various mutant states during interaction with small molecule variants of this drug class

could prove illuminating with regard to the problems of B-Raf hyperactivation and mutational tolerance more generally. Three B-Raf inhibitors are currently in use. These include sorafenib, a potent hyperactivation inducer in normal cells, [20](Zhang et al., 2015)[20][20][20][20][20] (Zhang et al., 2015)[20][20][20](Zhang et al., 2015)as well as dabrafenib and vemurafenib, two more modern inhibitors that also show potential problems with hyperactivation [14,20,21]. PLX7904, a modern hyperactivation paradox breaker is currently in preclinical development [20] while its analogue PLX8394 is in clinical trials. Here, we employ machine learning enhanced comparative analyses of MD simulation [22,23] to identify functionally conserved loop dynamics connected to normal ATP binding. We examine how the V600E mutation breaks the natural regulatory loop dynamics of the B-Raf protein. We compare the ability of each of these four B-Raf inhibitors (see Appendix A) to mimic functionally conserved loop dynamics (i.e. creating functional dynamic mimicry (FDM) in their B-Raf targets), and we quantify the impacts of known sensitizing and drug resistant mutations on these dynamics. We also examine the V600E sensitivity of loop dynamics in two ancient phylogenetically reconstructed protein models, so as to determine what evolutionary process may have led to unusual functional vulnerability of this important signaling protein to the V600E mutation.

Results

Functionally conserved B-Raf loop dynamics upon ATP binding in wildtype background

The comparative analysis of wildtype B-Raf dynamics in both its ATP-bound and unbound forms (via DROIDS 3.0) revealed upon binding, a significant dampening of atom fluctuation (i.e. decreased rmsf or negative dFLUX defined by the KL divergence in rmsf) in the ATP binding pocket associated with highly pronounced increase in fluctuation in the region of the activation loop as well as upstream toward the C terminal approximately affecting residues 599-625 (Figure 2A-C). Shifts in dynamics were significant according to multiple test corrected Kolmogorov-Smirnov tests at four major regions (Figure 2D) including the binding pocket, the catalytic + activation loops and at two regions centered at residue 550 and 660. When color mapped to the ATP bound structure, these dynamics exhibit large separation between the activation loop and P-loop (i.e. open active conformation for the protein) (Figure 2C). When trained to recognize binding interaction dynamics and later deployed upon MD validation runs for ATP bound wildtype B-Raf, our maxDemon 1.0 stacked machine learning model identified strong canonical correlations in learning performance indicating conserved binding dynamics across the regions connecting the ATP binding pocket to both the tip of the P-loop and a large region surrounding the catalytic and activation loops (Figure 2E-F). High R values in these regions indicate that the learning algorithms identified dynamics characteristic of B-Raf's interaction with ATP over 65% of the total time of simulation, in spite of the thermal fluctuation that is always present. The plots of local learning performance of each algorithm in each of the two validation runs (Supplemental Figure 1) also indicates that characteristic functional B-Raf binding dynamics corresponded well with the known functional regions of the protein as well as the C terminal region approximately from residue 625-675.

Fig 2. Comparative molecular dynamic analysis of ATP binding interaction and subsequent machine learning derived Raf inhibitor impacts in normal wild type human B-Raf kinase domain (PDB ID = 1uwh). *(A) Average atom fluctuation over time (i.e. RMSF or FLUX) for two ensembles of 300 x 0.5ns molecular dynamic simulations of inhibitor bound and unbound structures are shown and statistically compared. (B) Signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) from each amino acid backbone are plotted and (C) shown color mapped on the protein structure with*

red indicating amplified fluctuation and blue indicating stabilized regions. Colored bars in B and circled regions in D indicate the P-loop (red), ATP binding pocket (blue), catalytic loop (green) and a highly flexible activation loop (yellow). Yellow arrows also indicate location of the activation loop in B, D and E highlighting a highly pronounced increase in flexibility of this region during ATP binding. P values from a Benjamini-Hochberg corrected Kolmogorov-Smirnov (KS) test are also shown in (D) indicating significant differences in dynamics between bound and unbound structures. (E-F) The machine learning model identifies functionally conserved (i.e. correlated) molecular dynamics that clearly connect the ATP binding pocket (blue arrow; ATP structure is tan) to the activation loop (yellow arrow), p loop (red arrow) and surrounding structures (E-F). Molecular dynamic impacts of the binding interaction of 4 different Raf inhibitors (sorafenib, vemurafenib, dabrafenib and PLX7904) are summarized over both (G) the entire protein and functionally conserved regions (supplemental figure 1). The local impacts on dynamics and learning performance profiles (supplemental figure 2) upon which the canonical correlation between the two validation sets is derived are also given (supplemental figure 3).

Supplemental Figure 1. *Raw local machine learning performance used to derive canonical correlations for detecting functionally conserved dynamics are shown for comparison in both (A) wildtype and (B) V600E mutant backgrounds. Two validation runs are plotted (red and blue) using all seven machine learners. The greater correlation in A compared to B can be seen, indicating the effect of V600E in destroying the functional dynamics of B-raf that is normally responsive to ATP binding.*

Effects of Raf inhibitors on conserved B-Raf loop dynamics in wildtype background

When we analyzed the impacts (i.e. relative entropy of the canonical correlations indicating conserved dynamics) of binding for four Raf inhibitors relative to the normal interaction of ATP with wildtype B-Raf (Figure 2G), we find sorafenib more closely mimics ATP in the magnitude of its dynamic impacts on conserved regions (Supplemental Figure 2-3, globally, locally resp.) and total protein (Figure 1G) than more modern inhibitors, vemurafenib, dabrafenib and PLX7904. The highest dynamic impact was in PLX7904, a potent 'breaker' of the 'paradox of hyperactivation', suggestive that unlike sorafenib, its binding to wildtype B-Raf may render B-Raf unrecognizable to the MAPK pathway. Comparative dynamic analysis of B-Raf bound to each of the four inhibitors within wildtype backgrounds (dFLUX profiles in Supplemental Figure 4) also indicates that sorafenib most closely mimics ATP via its interaction with the binding pocket and subsequent amplifications of dynamics of the activation loop (i.e. see similarity between Figure 1B and Supplemental Figure 4A). While the modern inhibitors have dFLUX profiles somewhat similar to that of ATP, they all differ in key respects. Vemurafenib and dabrafenib also seem to affect P loop dynamics in ways that are very different than ATP (position 460-468 in Supplemental Figure 4B and 4C), while vemurafenib and PLX7904 exhibits more global disruption and dampening of dynamics across the whole B-Raf binding domain (indicated by widespread negative dFLUX in Supplemental Figure 4B and 4D).

Supplemental Figure 2*. Impacts of four Raf inhibitor drug class variants on functionally conserved ATP binding dynamics in wildtype B-raf protein. Additional runs with just ATP binding impact is also shown for validation.*

***Supplemental Figure 3**. Local impacts of four Raf inhibitor drug class variants on functionally conserved ATP binding dynamics in B-raf protein compared in both (A) wildtype and (B) V600E mutant backgrounds. Additional runs with just ATP binding impact is also shown for validation.*

***Supplemental Figure 4.** Comparative molecular dynamic analysis of Raf inhibitor binding interactions in normal wild type human B-Raf kinase domain. Inhibitors are (A) sorafenib, (B) vemurafenib, (C) dabrafenib and (D) PLX7904 (PDB ID = 1uwh, 4rzv, 4xv2 and 4xv1). Signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) from each amino acid backbone are plotted (left) shown color mapped on the protein structure (right) with red indicating amplified fluctuation and blue indicating stabilized regions.*

Malfunction of conserved B-Raf loop dynamics during ATP binding in V600E mutant

Comparative dynamics analysis of ATP binding interaction in the presence of the V600E mutation exhibited clear and striking differences in dynamics compared to the wildtype background. ATP binding not only stabilized the regions around the P-loop and ATP binding pocket, but also massively and significantly stabilized the dynamics of the region of the catalytic + activation loop and a large region from residue 650-680 (Figure 3 A-D) serving to freeze the activation loop into an open configuration characteristic of the V600E mutant (Figure 3C). In this genetic background, our machine learning model fails to identify any functionally conserved dynamics of ATP binding, except in a very small part of the region (650-680) that freezes to hold the activation loop (Figure 3E-F). The impacts of the four B-Raf inhibitors on protein dynamics were roughly equivalent to that of ATP in the mutant background (Figure 3G), but generally stronger than the impact of ATP and sorafenib in the wildtype background (Figure 2G). The local machine learning performance was generally poorer across the whole domain (i.e. closer to 0.5 in Supplemental Figure 1B) indicating that dynamics played a far less important functional role in the V600E mutant background than in wildtype. This is indicative of how V600E destroys the functional dynamics of ATP binding in B-Raf.

***Fig 3. Comparative molecular dynamic analysis of ATP binding interaction and subsequent machine learning derived Raf inhibitor impacts in mutant V600E human B-Raf kinase domain (PDB ID = 1uwh).** (A) Average atom fluctuation over time (i.e. RMSF or FLUX) for two ensembles of 300 x 0.5ns molecular dynamic simulations of inhibitor bound and unbound structures are shown and statistically compared. (B) Signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) from each amino acid backbone are plotted and (C) shown color mapped on the protein structure with red indicating amplified fluctuation and blue indicating stabilized regions. Colored bars in B and circled regions in D indicate the P-loop (red), ATP binding pocket (blue), catalytic loop (green) and a highly flexible activation loop (yellow). Yellow arrows also indicate location of the activation loop in B, D and E highlighting a highly pronounced increase in flexibility of this region during ATP binding. P values from a Benjamini-Hochberg corrected Kolmogorov-Smirnov (KS) test are also shown in (D) indicating significant differences in dynamics between bound and unbound structures. (E-F) The machine learning model fails to identify functionally conserved (i.e. correlated) molecular dynamics connecting the ATP binding pocket (blue arrow; ATP structure is tan) to the activation loop (yellow arrow), p loop (red arrow) and surrounding structures (E-F). Molecular dynamic impacts of the binding interaction of 4 different Raf inhibitors (sorafenib, vemurafenib, dabrafenib and PLX7904) are summarized over (G) the entire protein.*

Effects of Raf inhibitors on mutant B-Raf loop dynamics

Comparative dynamics analysis of the four B-Raf inhibitors in the V600E mutants (Supplemental Figure 5) indicated that sorafenib globally destabilized B-Raf upon interaction with the ATP binding pocket (Supplemental Figure 5A). This effect was larger toward the C terminal regions of the protein. In the mutant background, Vemurafenib appeared to mimic the functional effect of ATP binding similar to wildtype, but too a much larger degree, confirming a potential indication of its well-known selectivity to mutant B-Raf (Supplemental Figure 5B). Upon its interaction with the binding pocket, Dabrafenib induced strong stabilization in the area of the catalytic and activation loops, locking B-Raf into its open confirmation and perhaps enhancing its well-known effect as an ATP competitor (Supplemental Figure 5C). Alternatively, upon binding the mutant, PLX7904 exhibited large destabilizations of the P-loop, ATP binding pocket and the regions extending from the activation loop and the region of residue 650-680 that freezes the open confirmation in the absence of the drug (Supplemental Figure 5D). Close-up images of the impacts on B-Raf inhibitor binding on conserved protein dynamics in the loop regions are shown in Supplemental Figure 6. We note that PLX7904, as the paradox breaking inhibitor, shows substantially less activity near the V600 region and activating segment in general, as it was constructed to disrupt overall B-Raf structure and affect dimerization, particularly targeting the L505 position, in order to affect the helix there. In our study, we find that the region near 490-505 appear to be selectively affected by disruptions to dynamics, suggesting a dynamic mechanism for its previously observed hyperactivation paradox breaking behavior.

Supplemental Figure 5. *Comparative molecular dynamic analysis of Raf inhibitor binding interactions in mutant V600E human B-Raf kinase domain. Inhibitors are (A) sorafenib, (B) vemurafenib, (C) dabrafenib and (D) PLX7904 (PDB ID = 1uwh, 4rzv, 4xv2 and 4xv1). Signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) from each amino acid backbone are plotted (left) shown color mapped on the protein structure (right) with red indicating amplified fluctuation and blue indicating stabilized regions.*

Supplemental Figure 6*. Close-up images of the mapping of conserved protein dynamics (darker regions) affected under binding of four raf inhibitors; (A) sorafenib, (B) PLX7904, (C) vemurafenib, and (D) dabrafenib. Arrow indicate ATP binding pocket (blue), activation loop (yellow) and P loop (red).*

Biphasic behavior in dynamics of B-Raf activation loop

A series of shorter comparative MD production runs were averaged to collect 'snapshots' in time and to investigate the presence of quasi-stable dynamic states. In most cases, the dFLUX (i.e. KL divergence) profiles in these runs matched that of longer runs. However, we did occasionally encounter a less common dFLUX profile around the V600 site, indicating strongly biphasic dynamic behavior in this region. In this situation, the typical combination of dampened and normal amplification of atom fluctuation (Supplemental Figure 7A) is replaced by only strong amplification at the same site (Supplemental Figure 7B). This potentially indicated the presences of very complex biphasic, or 'switch-like' behavior of the dynamics in the region where V600E sensitivity occurs.

Supplemental Figure 7. *Natural biphasic behavior of the activation loop captured by comparative molecular dynamic analysis of PLX7904 binding interaction in mutant V600E B-Raf kinase domain (PDB*

ID = 4xv1). Colored circled regions indicate the P-loop (red), ATP binding pocket (blue), catalytic loop (green) and a highly flexible activation loop (yellow). The activation loop is shown in two stable states, (A) open conformation with and highly flexible activation loop, and (B) closed conformation with highly dampened activation loop. NOTE: simulation runs were averaged over only 20x0.1ns production runs (i.e. fewer sampling runs taken over shorter time periods than training sets in Figure 2 and 3).

Comparison of impacts of drug sensitivity and resistant mutations on B-Raf loop dynamics

To further investigate the effect of both genetic and drug class variation on functional B-raf loop dynamics, we compared the dynamic impacts on B-Raf for a variety of drug sensitizing and drug resistance linked mutations for each of the four Raf inhibitors, sorafenib, vemurafenib, dabrafenib, and PLX7904 (Supplemental Table 1, Figures 4-7). During the interaction of hyperactivation-inducing sorafenib with the ATP binding pocket, dynamic shifts in the activation loop are clearly evident (as large positive and negative dFLUX values in Figure 4A) which are quite similar in effect to changes in this region caused by ATP binding (Figure 2A). Functional dynamic shifts identified by our machine learning model are also quite evident upon sorafenib binding (Figure 4B-C) and similarly connect the ATP binding pocket to the outer bend of the activation loop when ATP binds (Figure 2B). The V600E mutation which sensitizes sofafenib targeting to B-Raf, was found to greatly impact loop dynamics when compared to validation runs with no mutation (Figure 4D) and corroborating its large impact on functional dynamics in general (Figure 3). The three remaining drug resistant mutations (G464V, G469E, G469V) show less impact, but are collectively larger than validation, suggesting that mutations disrupting dynamics of the P loop are key in creating drug resistance to sorafenib (Figure 4D). In stark contrast to the sorafenib results, the hyperactivation-breaking compound, PLX7904, while also causing similar dynamics shifts to the activation loop (Figure 5A), fails to produce hardly any dynamic changes that are identified as functional by our machine learning classification model (Figure 5B-C). Thus, it appears that while PLX7904 is a potent competitive binder to mutant B-Raf, it also has the benefit of potentially disrupting any functional dynamics that could allow stable interactions with other components of the MAPK pathway. Our analysis of mutational impacts on functional dynamics (Figure 4D) also support this, and suggest that changes in PLX7904 efficacy caused by mutation are likely independent of the normal evolutionary conserved function of B-Raf. Analysis of vemurafenib (Figure 6) shows marked similarity to sorafenib, suggesting that it is also prone to mimic the functional dynamics of ATP binding (Figure 6B-C) but with softening of the P-loop as well (Figure 6A). A survey of 18 drug impacting mutations does not show any clear trend except that mutational impacts are largest in regions affecting either the P-loop or the activation loop (Figure 6D). Similar analysis of dabrafenib indicates a hardening of P-loop dynamics (Figure 7A) and a pattern of less coordinated impact on conserved dynamics that fails to connect the ATP binding pocket to the activation loop (Figure 7B-C), supporting our hypothesis that more modern Raf inhibitors are less like ATP in their overall effect on functional shifts in B-Raf dynamics during binding. Mutations affecting dabrafenib function also appear to be largely confined to P-Loop and activation loop, with the exception of L514V, a site whose side-chain actively engages ATP during normal B-Raf function.

Table 1. *Genetic variants affecting drug activity of four Raf inhibitors (sorafenib, vemurafenib, dabrafenib, PLX7904).*

361 ***Fig 4. Comparative molecular dynamic analysis of hyperactivation-inducing sorafenib binding in***
362 ***wildtype BRAF and subsequent mutational impacts on conserved BRAF dynamics.*** *Images represent*
363 *the (A) signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) color-*
364 *mapped to the BRAF structure with red indicating amplified fluctuation and blue indicating stabilized*
365 *regions and (B) color-mapped functionally conserved dynamic regions (dark gray) identified by (C)*
366 *significant canonical correlation of performance in machine learning validation runs. (D) The impacts of*
367 *four known drug sensitizing mutations (V600E, G469V, G469E, G464V) are also compared. NOTE: blue,*
368 *yellow and red arrows indicate drug binding site, activation loop and p loop (resp.)*

369 ***Fig 5. Comparative molecular dynamic analysis of hyperactivation-breaking PLX7904 binding in***
370 ***wildtype BRAF and subsequent mutational impacts on conserved BRAF dynamics.*** *Images represent the*
371 *(A) signed symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) color-*
372 *mapped to the BRAF structure with red indicating amplified fluctuation and blue indicating stabilized*
373 *regions and (B) color-mapped functionally conserved dynamic regions (dark gray) identified by (C)*
374 *significant canonical correlation of performance in machine learning validation runs. (D) The impacts of*
375 *three known drug sensitizing mutations (V600E, V600R, V600D) and one drug resistant mutation (G466E)*
376 *are also compared. NOTE: blue, yellow and red arrows indicate drug binding site, activation loop and p*
377 *loop (resp.)*

378 ***Fig 6. Comparative molecular dynamic analysis of vemurafenib binding in wildtype BRAF and***
379 ***subsequent mutational impacts on conserved BRAF dynamics.*** *Images represent the (A) signed*
380 *symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) color-mapped to the*
381 *BRAF structure with red indicating amplified fluctuation and blue indicating stabilized regions and (B)*
382 *color-mapped functionally conserved dynamic regions (dark gray) identified by (C) significant canonical*
383 *correlation of performance in machine learning validation runs. (D) The impacts of 18 known drug*
384 *sensitizing and resistant mutations are also compared. See Table 1 for classifications. NOTE: blue, yellow*
385 *and red arrows indicate drug binding site, activation loop and p loop (resp.)*

386 ***Fig 7. Comparative molecular dynamic analysis of dabrafenib binding in wildtype BRAF and***
387 ***subsequent mutational impacts on conserved BRAF dynamics.*** *Images represent the (A) signed*
388 *symmetric Kullback-Leibler (KL) divergences in local RMSF distributions (i.e. dFLUX) color-mapped to the*
389 *BRAF structure with red indicating amplified fluctuation and blue indicating stabilized regions and (B)*
390 *color-mapped functionally conserved dynamic regions (dark gray) identified by (C) significant canonical*
391 *correlation of performance in machine learning validation runs. (D) The impacts of 10 known drug*
392 *sensitizing and resistant mutations are also compared. See Table 1 for classifications. NOTE: blue, yellow*
393 *and red arrows indicate drug binding site, activation loop and p loop (resp.)*

394

395 **Phylogenetic analysis of V600E sensitivity regarding B-Raf loop dynamics**

396 A minimum evolution tree (Figure 8A) and maximum likelihood tree (supplemental Figure 8) was used to
397 identify protein level variants for tetrapod (= human), ray-finned fish, and insect. The wildtype genetic
398 backgrounds incorporating 9 amino acid replacements (fish) and 95 replacements (insect) show no
399 elevated dynamic impacts over the validation run (wildtype tetrapod). However, V600E/A600E mutants
400 both demonstrate elevated disruption of conserved ATP binding dynamics (Figure 8B) indicating that B-
401 Raf sensitivity to this mutation has an ancient origin. The very large increase in dynamic impact of V600E

402 in ray-finned fish over that of A600E in insects indicates a large increase in V600E sensitivity in early
403 jawed vertebrate evolution as well (Figure 8B). Local mapping of these impacts (Figure 8C) indicates
404 involvement of all three functional regions (P-loop at position 13-25, ATP binding pocket at position 60-
405 100, and activation loop at position 153-190) in the functional evolution of this sensitivity. The dynamic
406 impacts of V600E and A600E in these ancient backgrounds are also shown color-mapped to the B-Raf
407 structure (Figure 8D and 8E resp.).

408 ***Fig 8. Comparative molecular dynamic analysis of ATP binding in wildtype BRAF orthologs.*** *Orthology*
409 *was determined via phylogenetic analysis to derive a (A) minimum evolution tree bootstrapped 500*
410 *times using the Kimura 2 parameter model with Gamma correction for multiple hits and invariant sites.*
411 *This best model was determined via multi-model inference (BIC, AICc). (B-C) With a goal of determining*
412 *when V600E sensitivity first arose in this phylogeny, we compare the conserved region impacts of ATP*
413 *binding on BRAF dynamics in wildtype and V600E mutated genetic backgrounds of ray-finned fish and*
414 *insect ortholog models after training on the functional ATP binding in wildtype tetrapod (i.e. human)*
415 *background. These impacts are shown both (B) globally and (C) locally, where regions of low mutational*
416 *tolerance are observed corresponding to the P-loop, ATP binding pocket, and activation loop and*
417 *adjacent downstream residues (shaded yellow). The impacts of the genetic variation of each variant is*
418 *color mapped to the wildtype and mutant backgrounds in both (D) fish and (E) insect. NOTE: in the insect*
419 *ortholog, this mutation is A600E rather than V600E. A maximum likelihood tree is also shown in*
420 *Supplemental Figure 7.*

421 ***Supplemental Figure 8.*** *A maximum likelihood tree bootstrapped 500 times using the Kimura 2*
422 *parameter model with Gamma correction for multiple hits and invariant sites. This best model was*
423 *determined via multi-model inference (BIC, AICc). Insect outgroup is not shown to enhance readability*
424 *of the bootstrap values.*

425

426 Discussion

427 Given their involvement in a variety of cellular processes, protein kinases, specifically serine/threonine
428 kinases play a role in all of the major cancers. Many serine/threonine kinases have altered expression in
429 tumors [24]. Given that this class of enzyme is a proven putative and attractive target for the
430 development of cancer therapies, our study of the function and evolution of the molecular dynamics of
431 the serine/threonine kinase B-Raf has the potential to guide future studies and experiments aimed at
432 elucidating efficacy and rationale design of therapeutics. Through our novel high performance
433 computing approach to machine learning applied to comparative functional molecular dynamics
434 simulations, we have demonstrated that ATP binding to the CR3 domain of B-Raf protein induces
435 evolutionarily functionally conserved changes in dynamics that physically connect ATP binding pocket
436 with both the P-loop and activation segment. The region just downstream of the mutation intolerant
437 valine codon 600 is especially affected in its dynamics by ATP binding remotely on the structure of the
438 protein KD. The V600E mutation massively disrupts these evolutionary conserved shifts in dynamics that
439 are obviously a key complex functional feature of BRAF signaling and MAPK regulation. Hyperactivation
440 inducing drugs, especially the first generation drug sorafenib, appears to mimic functionally conserved
441 dynamics of natural ATP binding function in the CR3 domain, inducing almost the same spatial pattern of
442 functionally conserved protein dynamics in our simulations as ATP itself. Vemurafenib and dabrafenib

show less of this tendency, but still induce functional dynamics in the region joining the ATP binding pocket to the P-loop and activation loop segment of B-Raf. In contrast, the hyperactivation breaking drug PLX7904 shows almost no such propensity for mimicry of dynamics of the ATP agonist. See close-up comparison of the functional dynamic mimicry (FDM) of ATP caused by the four Raf inhibitors in Supplemental Figure 6. Our survey of single amino acid replacements most affecting patient drug sensitivity or resistance to the four inhibitors indicates that mutations impacting the three hyperactivation inducing drugs (i.e. sorefenib, vemurafenib and dabrafenib) appear to target functionally conserved aspects of P-loop and activation loop dynamics. This might be expected under a somatic mutation-selection regime induced by therapeutic treatments of tumors where functional B-Raf dynamics are still potentially in play, albeit likely subverted into a malfunctional state from the standpoint of the patient. However, PLX7904 currently has very few known sensitizing or resistant mutations, an observation that may reflect lack of experimentation with the new compound, but may also prove consistent with our demonstration that it largely avoids inducing the same sort of dynamic changes as ATP upon binding the same location. Finally, our molecular evolutionary study, indicates that B-Raf sensitivity to V600E has a likely ancient origin, perhaps as old as the Raf kinase protein family itself. However, the sensitivity of functionally conserved B-Raf dynamics to V600E, appears to have greatly increased sometime during early jawed vertebrate evolution, as represented by extant sequences across all modern ray-finned fishes. As a modulator of multicellular growth, B-Raf likely has origins in early eukaryotes, however poor alignment quality to putative orthologs in *C. elegans* and *S. cerevisiae* prevented us from exploring this possibility further.

Conclusion

We propose that functional dynamic mimicry (FDM), as observed here in B-Raf inhibitors, is an important risk factor to consider in future rational drug design. This an especially important consideration when targeting sites of natural agonists involved in signal activation in pathways like MAPK which are typically downregulated after early embryonic development, and a subsequently problematic in all cancers. Pharmaceuticals that avoid FDM while effectively targeting and binding active sites of proteins in these pathways, are most likely disrupting the functional dynamics of protein domains to the extent that they are no longer recognized by the pathways in which they normally participate. In the case of pathways downregulated after early development, this affect is ideal, as it destroys signal transduction. Our DROIDS 3.0 + maxDemon 1.0 computational software [22,23] can allow for inexpensive screening against unintended effects of FDM, potentially producing better patient products while saving pharmaceutical industry wasted time, effort and money in preclinical development and clinical trial. We also conclude that a more evolutionary and biophysical perspective around complex problems like FDM that are potentially linked to naturally evolved tolerances to protein mutation are greatly needed in future biomedical research. This cannot be easily addressed by the combination of DNA sequence-based bioinformatics and cell/animal-based experimental research alone. Proteins are nanometer sized machines operating in the complex context of soft matter dynamics. They are often imperfect in their interactions with other proteins and incompletely optimized in terms of their functional evolution. The development of modern drug therapies could benefit by embracing a more functional evolutionary perspective that is grounded in computational biophysical method.

Appendix A

Supplementary information on the four inhibitors: The four B-Raf inhibitors (BRAFi) discussed in this manuscript are all small molecule compounds that are FDA approved for use: 1. in mutated BRAF in metastatic melanoma (Vemurafinib, Dabrafenib) or 2. in other carcinomas (Sorafenib, for advanced renal cell, thyroid, and hepatocellular carcinomas) as a more general kinase inhibitor (active against BRAF, VEGFR, etc). PLX7904 is in preclinical development.

Sorafenib (also, Nexavar, BAY 43-9006, MW: 464.8 g/mol): A member of the class of phenylureas that acts as a kinase inhibitor and angiogenesis inhibitor. It is nonspecific, and does not distinguish between WT and mutant B-Raf forms, and has higher specificity for C-Raf compared to B-Raf. It is known to potently induce hyperactivation, and leads to rapid development of drug resistance.

Vemurafinib (also, PLX4032; Zelboraf; MW: 489.9 g/mol): An ATP-competitive kinase inhibitor selective for B-Raf V600E. It functions to decrease signaling through the MAPK pathway by inhibiting the kinase domain of mutant BRAF, but is known to also induce hyperactivation in WT BRAF. Shows increased specificity for B-Raf V600E relative to B-Raf WT or C-Raf.

Dabrafenib (also, Tafinlar, GSK2118436; 519.6 g/mol): An ATP-competitive kinase inhibitor selective for B-Raf V600 (E/K/R) mutations, which acts to induce inhibition of phosphorylated extracellular signal-regulated kinase (ERK). Often given in combination with MEK inhibitor Trametinib to achieve a more favorable resistance profile.

PLX7904 (also PB04; MW: 512.5 g/mol): A selective B-Raf kinase inhibitor known to be a potent paradox-breaker. able to efficiently inhibit activation of ERK1/2 in mutant BRAF melanoma cells but does not hyperactivate ERK1/2 in mutant RAS-expressing cells. Was designed to interact with position 505 in B-Raf, shifting the Raf dimer interface, and cause inhibition via long-range structural alterations. Along with analogue PLX8394, is in development as BRAFi capable of uncoupling RAF blocking and MAPK pathway activation.

Models

PDB structure preparation and ligand force field modification

Structures of the ATP binding domain of B-Raf bound to various Raf inhibitors were obtained from the Protein Data Bank (PDB). These were PDB ID 1uwh (wildtype B-Raf bound to sorafenib), 4rzv (vemurafenib bound to V600E mutant), 4xv1 (PLX7904 bound to V600E mutant), 4xv2 (dabrafenib bound to V600E mutant). Loop modeling and refinement were conducted on the activation loop where needed using Modeller in UCSF Chimera [25]. Complementary mutant (V600E in 1uwh) and wildtype structures (E600V in 4rzv, 4xv1,4xv2) were computationally induced in UCSF Chimera 1.13 [25] by first swapping amino acids using optimal configurations in the Dunbrack rotamer library, then using 2000 steepest gradient descent steps of energy minimization to relax the regions around the amino acid replacement. To obtain ATP bound versions of wildtype B-raf and the V600E mutant, the sorafenib inhibitor was deleted from 1uwh and 1uwj, then ATP was docked using the AutoDock Vina program [26] after reducing both structures (adding H). ATP docking positions most closely overlapping sorafenib were chosen. Unbound forms of all the structures were obtained by deleting the ligands in the bound structure and performing energy minimization for 5000 steps. Prior to molecular dynamic simulation in

523 Amber18, the GAFF2 force field modifications for ATP and the small molecule ligands were generated
524 using sqm, a semiempirical and DFTB quantum chemistry program employed by the Antechamber
525 (Ambertools18) software suite [27,28].

526 **MD simulation protocols**

527 Large ensembles of graphic processing unit (GPU) accelerated molecular dynamic simulations were
528 prepared and conducted using the particle mesh Ewald method employed by pmemd.cuda in Amber18
529 [29,30] via the DROIDS v3.0 interface (Detecting Relative Outlier Impacts in Dynamic Simulation) [22,23].
530 Simulations were run on a Linux Mint 19 operating system mounting two Nvidia Titan Xp graphics
531 processors. Explicitly solvated protein systems were prepared using tLeAP (Ambertools18) using the
532 ff14SB protein force field [31] in conjunction with modified GAFF2 small molecule force field [28].
533 Solvation was generated using the Tip3P water model [32] in a 12nm octahedral water box and
534 subsequent charge neutralization with Na+ and Cl- ions. After energy minimization, heating to 300K, and
535 10ns equilibration, an ensemble of 200 MD production runs each lasting 0.5 ns of time were created for
536 both ligand bound and unbound B-Raf. Each MD production run was preceded by a single random
537 length short spacing run selected from a range of 0 to 0.25ns to mitigate the effect of chaos (i.e.
538 sensitively to initial conditions) in the driving ensemble differences in the MD production runs [22]. All
539 MD was conducted using an Andersen thermostat [33] under constant pressure. Root mean square
540 atom fluctuations (rmsf) were calculated using the atomicfluct function in CPPTRAJ [34].

541 **Comparative dynamics of bound and unbound functional states**

542 The symmetric Kullback-Leibler (KL) divergence or relative entropy between the distributions of atom
543 fluctuation (i.e. root mean square fluctuation or rmsf taken from 0.01 ns time slices of total MD
544 simulation time) on ligand bound and unbound B-Raf were computed using DROIDS and color mapped
545 to the protein backbone with individual amino acid resolution to the bound structures using a
546 temperature scale (i.e. where red is hot or amplified fluctuation and blue is cold or dampened
547 fluctuation). Details of this calculation and representation can be found in our previous publication [22].
548 The reference state of the protein is unbound while the query state is bound. Therefore, this pairwise
549 comparison is used to represent the functional impact of ligand binding on the protein's normal
550 unbound motion, where it is expected that ligand contacts would typically dampen the fluctuation of
551 atoms around the ATP binding pocket to some degree. As ATP binding acts to destabilize the activation
552 loop and P-loop regions of B-Raf, this analysis will allow quantification of dynamic impacts that amplify
553 loop motions as well.

554 **Machine learning performance and detection of conserved dynamic function**

555 We applied machine learning analysis to examine normal ATP binding in wildtype B-Raf. We then
556 compare these results to a similar analysis of ATP binding in V600E mutants to ascertain how this
557 mutation alters functionally conserved dynamics in the B-Raf protein system. The detection of
558 functionally conserved protein dynamics in wildtype B-Raf followed the machine learning-based method
559 outlined in our previous DROIDS 3.0 + maxDemon 1.0 software papers [22,23]. In summary, a stacked
560 machine learning model implementing seven classification methods (K nearest neighbors, naïve Bayes,
561 linear/quadratic discriminant analysis, support vector machine, random forest and adaptive boosting)
562 was individually trained upon each amino acid's fluctuation data that was pre-classified to the ligand
563 bound and the unbound dynamic states. These learners were deployed upon the identical amino acids

within two new 5ns MD simulations of the ligand bound structure and classifications of atom fluctuation (i.e. rmsf calculated for 500 x 0.01 ns time slices) where 0 = unbound and 1 = bound states were averaged to obtain a learning performance value for each amino acid. Average learning performance profiles (i.e. accuracy and precision) are generated along the protein from N to C terminus. Note that a value of 0.5 at a given position would indicate that the learning model could not differentiate between bound and unbound dynamics at that residue. Local significant canonical correlations in learning performance between the two validation runs on the ligand bound structure indicate where any position-dependent molecular dynamics that is functionally conserved with regards to the binding interaction has occurred. The correlation was calculated within a 20 residue sliding window along the learning profile.

$$conserved_{dynamics} = significant\ (CC_{self})$$

The central concept here is that if any atom motion is specific to a given functional state (i.e. in this case protein bound to ligand, vs unbound protein), then after training, the machine learners will be able to successfully classify this motion when it is observed in subsequent identical validation MD runs. When the learners are able to repeatedly identify functional state dynamics that locally associate with a specific position on the protein, this will create correlated profiles in learning performance. As our machine learning model is a stacked model, incorporating seven learning algorithms, it is more robust to artifacts related to any single method, and the resulting correlation is canonical as well (see Supplemental Figure 1 for an example of correlated and uncorrelated sets of learning profiles). We can label the dynamics as 'functionally conserved' if the machine learning model can consistently identify the functional dynamic states upon which it trained across multiple and independent validation runs. Two or more identically prepared validation runs are used to examine whether statistically significant canonical correlations (i.e. conserved dynamics) are locally identifiable by a significant local Wilk's lambda value for a local canonical correlation coefficient.

Method of genetic and drug variant impact assessment

The local impacts of mutations and drug class variants were obtained by first deriving the canonical correlation of the learner performance profile of a given variant to the wildtype B-Raf in its normal functionally bound state (= $CC_{variant}$). This also uses a 20 residue sliding window. The impact metric is generated by comparing canonical correlation of the variant ($CC_{variant}$) to the self-similar dynamic correlation (CC_{self}) defining functionally conserved dynamics above, using the common definition of relative entropy below

$$variant_{impact} = CC_{self} * log\frac{CC_{variant}}{CC_{self}}$$

Effectively, this impact metric captures the difference in conserved dynamics between the wildtype ligand bound state (here ATP or one of the four Raf inhibitors) and the variant states of binding created by genetic mutation or different drug ligand molecules. This occurs on the same scale as that used to differentiate bound from unbound states during the earlier training of the stacked learning model. More details regarding our methodological approach to identifying functionally conserved dynamics and subsequent impacts of genetic and drug class variants can be obtained in our prior publication [22,23].

The relative impacts of four B-Raf inhibitors, sorafenib, dabrafenib, vemurafenib and PLX7904 were assessed after initially training the stacked machine learning model on ATP bound and unbound B-

604 Raf within both the wildtype and the V600E mutant genetic background. Several drug affecting genetic
605 mutants were also examined after training the learning model on normal functional binding interaction
606 of each drug in the wildtype genetic background. These drug resistant and drug sensitizing mutations
607 were identified using the Variant Interpretation for Cancer Consortium (VICC) Meta-Knowledgebase
608 [35]. Variants were selected that contained literature support for the functional impact on at least one
609 of the four B-Raf inhibitors (Supplemental Table 1).

610 **Phylogenetic analysis of B-Raf sensitivity to V600E**

611 A set of 27 BRAF ortholog sequences were obtained using the National Center for Biotechnology
612 Information (NCBI) ortholog finder tool. These sequences were aligned using MEGA 7.0 [36] Clustal
613 multiple sequence alignment tool and then manually trimmed to contain only the ATP binding domain.
614 Phylogenetic model selection was also conducted in MEGA to identify the most likely (best fitting)
615 evolutionary substitution model. This was identified via both Bayesian and Akaike Information Criterion
616 to be the Kimura 2-parameter model with extra terms for both gamma correction for multiple same site
617 substitution and correction for invariants sites (i.e. K2+G+I). A minimum evolution tree and a maximum
618 likelihood tree bootstrapped 500 times confirmed the same tree topology with high confidence. Amino
619 acid replacements distributed on this topology identified three protein variants, one each for tetrapods,
620 ray-finned fish and insects. Amino acids differences from tetrapods identified 9 amino acid replacements
621 since a common ancestor with ray-finned fish and 95 replacements since a common ancestor with
622 insects. Structural models for the two ancestral protein variants were manually generated in UCSF
623 Chimera using swapaa commands to find non-clashing Dunbrack library rotamer configurations. These
624 models were then subsequently relaxed with 2000 steps of energy minimization via steepest gradient
625 descent. A V600E variant was subsequently created from the wildtype ray-finned fish model and an
626 A600E variant was created from the wildtype insect model. DROIDS software was used to train upon
627 ATP binding interaction in the wildtype tetrapod model (comparing ATP bound and unbound dynamics
628 as described above). DROIDS 3.0 and maxDemon 1.0 machine learning analysis (described above) was
629 then used to identify functionally conserved dynamics in tetrapod B-Raf and subsequently compare
630 mutational impacts on B-Raf dynamics for both the ancestral wildtype and mutant backgrounds in ray-
631 finned fish and insects.

632 The author's software can be accessed at the promotional website and GitHub repository links below

633 https://people.rit.edu/gabsbi/

634 https://github.com/gbabbitt/DROIDS-3.0-comparative-protein-dynamics

635 Acknowledgements

636 We acknowledge Nvidia corporation for a GPU hardware grant, and the College of Science at the
637 Rochester Institute of Technology for internal grant support.

638

639 List of abbreviations

640 FDM - functional dynamic mimicry

641 MD - molecular dynamic (simulation)

642 GPU – graphic processor unit

643 PDB – Protein Data Bank

644 NCBI – National Center for Biotechnology Information

645 BRAF – B-Raf gene sequence

646 B-Raf – protein product of BRAF gene

647 ATP – adenosine triphosphate

648 DNA – deoxyribonucleic acid

649 MAPK – mitogen-activated protein kinase

650 BRAFi – B-Raf inhibitor

651 V600E – valine to glutamic acid mutation and position 600

652 CR1 CR2 CR3 – conserved region 1, 2, or 3

653 KD – kinase domain

654 P-loop – ATP stabilizing loop

655 KL divergence – Kullback-Leibler divergence

656 dFLUX – change in atom fluctuation or rmsf

657 rmsf – root mean square fluctuation

658 CC – canonical correlation

659 DROIDS – software acronym for 'detecting outlier impacts in dynamic simulation', our software package
660 for comparative protein dynamics simulation

661 maxDemon – software acronym for 'Maxwell's demon', our machine learning application for DROIDS

662

663 References

664 1. Andersen BL, Shapiro CL, Farrar WB, Crespin T, Wells-DiGregorio S. Psychological Responses to
665 Cancer Recurrence. Cancer. 2005;104: 1540–1547. doi:10.1002/cncr.21309

666 2. Yang H-C, Thornton LM, Shapiro CL, Andersen BL. Surviving Recurrence: Psychological and Quality-
667 of-life Recovery. Cancer. 2008;112: 1178–1187. doi:10.1002/cncr.23272

668 3. Janz NK, Li Y, Zikmund-Fisher BJ, Jagsi R, Kurian AW, An LC, et al. The Impact of Doctor-Patient
669 Communication on Patients' Perceptions of their Risk of Breast Cancer Recurrence. Breast Cancer
670 Res Treat. 2017;161: 525–535. doi:10.1007/s10549-016-4076-5

671 4. Brown A-L, Li M, Goncearenco A, Panchenko AR. Finding driver mutations in cancer: Elucidating
672 the role of background mutational processes. PLOS Comput Biol. 2019;15: e1006981.
673 doi:10.1371/journal.pcbi.1006981

674 5. Gatenby RA, Zhang J, Brown JS. First Strike–Second Strike Strategies in Metastatic Cancer: Lessons
675 from the Evolutionary Dynamics of Extinction. Cancer Res. 2019 [cited 4 Sep 2019].
676 doi:10.1158/0008-5472.CAN-19-0807

677 6. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment
678 of metastatic castrate-resistant prostate cancer. Nat Commun. 2017;8: 1816. doi:10.1038/s41467-
679 017-01968-5

680 7. Bürger R. Mutation-Selection Models in Population Genetics and Evolutionary Game Theory. In:
681 Kurzhanski AB, Sigmund K, editors. Evolution and Control in Biological Systems. Springer
682 Netherlands; 1989. pp. 75–89.

683 8. Hartl DL. Mutation-selection balance with stochastic selection. Genetics. 1977;86: 687–696.

684 9. Nunney L. The population genetics of multistage carcinogenesis. Proc R Soc B Biol Sci. 2003;270:
685 1183–1191. doi:10.1098/rspb.2003.2351

686 10. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, et al. RAF inhibitors prime
687 wild-type RAF to activate the MAPK pathway and enhance growth. Nature. 2010;464: 431–435.
688 doi:10.1038/nature08833

689 11. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N. RAF inhibitors transactivate RAF dimers and
690 ERK signaling in cells with wild-type BRAF. Nature. 2010;464: 427–430. doi:10.1038/nature08902

691 12. Villanueva J, Vultur A, Herlyn M. Resistance to BRAF inhibitors: Unraveling mechanisms and future
692 treatment options. Cancer Res. 2011;71: 7137–7140. doi:10.1158/0008-5472.CAN-11-1243

693 13. Montagut C, Settleman J. Targeting the RAF-MEK-ERK pathway in cancer therapy. Cancer Lett.
694 2009;283: 125–134. doi:10.1016/j.canlet.2009.01.022

695 14. Wan PTC, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, et al. Mechanism of activation
696 of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell. 2004;116: 855–867.

697 15. Roskoski R. RAF protein-serine/threonine kinases: structure and regulation. Biochem Biophys Res
698 Commun. 2010;399: 313–317. doi:10.1016/j.bbrc.2010.07.092

699 16. Durrant DE, Morrison DK. Targeting the Raf kinases in human cancer: the Raf dimer dilemma. Br J
700 Cancer. 2018;118: 3–8. doi:10.1038/bjc.2017.399

701 17. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in
702 human cancer. Nature. 2002;417: 949–954. doi:10.1038/nature00766

703 18. Rezaei Adariani S, Buchholzer M, Akbarzadeh M, Nakhaei-Rad S, Dvorsky R, Ahmadian MR.
704 Structural snapshots of RAF kinase interactions. Biochem Soc Trans. 2018;46: 1393–1406.
705 doi:10.1042/BST20170528

706 19. Ascierto PA, Kirkwood JM, Grob J-J, Simeone E, Grimaldi AM, Maio M, et al. The role of BRAF V600
707 mutation in melanoma. J Transl Med. 2012;10: 85. doi:10.1186/1479-5876-10-85

708 20. Zhang C, Spevak W, Zhang Y, Burton EA, Ma Y, Habets G, et al. RAF inhibitors that evade
709 paradoxical MAPK pathway activation. Nature. 2015;526: 583–586. doi:10.1038/nature14982

710 21. Karoulia Z, Wu Y, Ahmed TA, Xin Q, Bollard J, Krepler C, et al. An Integrated Model of RAF Inhibitor
711 Action Predicts Inhibitor Activity against Oncogenic BRAF Signaling. Cancer Cell. 2016;30: 485–498.
712 doi:10.1016/j.ccell.2016.06.024

713 22. Babbitt GA, Mortensen JS, Coppola EE, Adams LE, Liao JK. DROIDS 1.20: A GUI-Based Pipeline
714 for GPU-Accelerated Comparative Protein Dynamics. Biophys J. 2018;114: 1009–1017.
715 doi:10.1016/j.bpj.2018.01.020

716 23. Babbitt GA, Fokoue EP, Evans JR, Diller KI, Adams LE. DROIDS 3.0 - detecting genetic and drug class
717 variant impact on conserved protein binding dynamics. Biophys J. 2019;0.
718 doi:10.1016/j.bpj.2019.12.008

719 24. Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, Nebuloni M, et al. Frequent alterations
720 in the expression of serine/threonine kinases in human cancers. Cancer Res. 2006;66: 8147–8154.
721 doi:10.1158/0008-5472.CAN-05-3489

722 25. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a
723 visualization system for exploratory research and analysis. J Comput Chem. 2004;25: 1605–1612.
724 doi:10.1002/jcc.20084

725 26. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring
726 function, efficient optimization and multithreading. J Comput Chem. 2010;31: 455–461.
727 doi:10.1002/jcc.21334

728 27. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in
729 molecular mechanical calculations. J Mol Graph Model. 2006;25: 247–260.
730 doi:10.1016/j.jmgm.2005.12.005

731 28. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber
732 force field. J Comput Chem. 2004;25: 1157–1174. doi:10.1002/jcc.20035

733 29. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular
734 simulation programs. J Comput Chem. 2005;26: 1668–1688. doi:10.1002/jcc.20290

735 30. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular
736 Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem
737 Theory Comput. 2013;9: 3878–3888. doi:10.1021/ct400314y

738 31. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the
739 Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput.
740 2015;11: 3696–3713. doi:10.1021/acs.jctc.5b00255

741 32. Mao Y, Zhang Y. Thermal conductivity, shear viscosity and specific heat of rigid water models.
742 Chem Phys Lett. 2012;542: 37–41. doi:10.1016/j.cplett.2012.05.044

743 33. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem
744 Phys. 1980;72: 2384–2393. doi:10.1063/1.439486

745 34. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular
746 Dynamics Trajectory Data. J Chem Theory Comput. 2013;9: 3084–3095. doi:10.1021/ct400341p

747 35. Wagner AH, Walsh B, Mayfield G, Tamborero D, Sonkin D, Krysiak K, et al. A harmonized meta-
748 knowledgebase of clinical interpretations of cancer genomic variants. bioRxiv. 2018; 366856.
749 doi:10.1101/366856

750 36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for
751 Bigger Datasets. Mol Biol Evol. 2016;33: 1870–1874. doi:10.1093/molbev/msw054

752

753

754

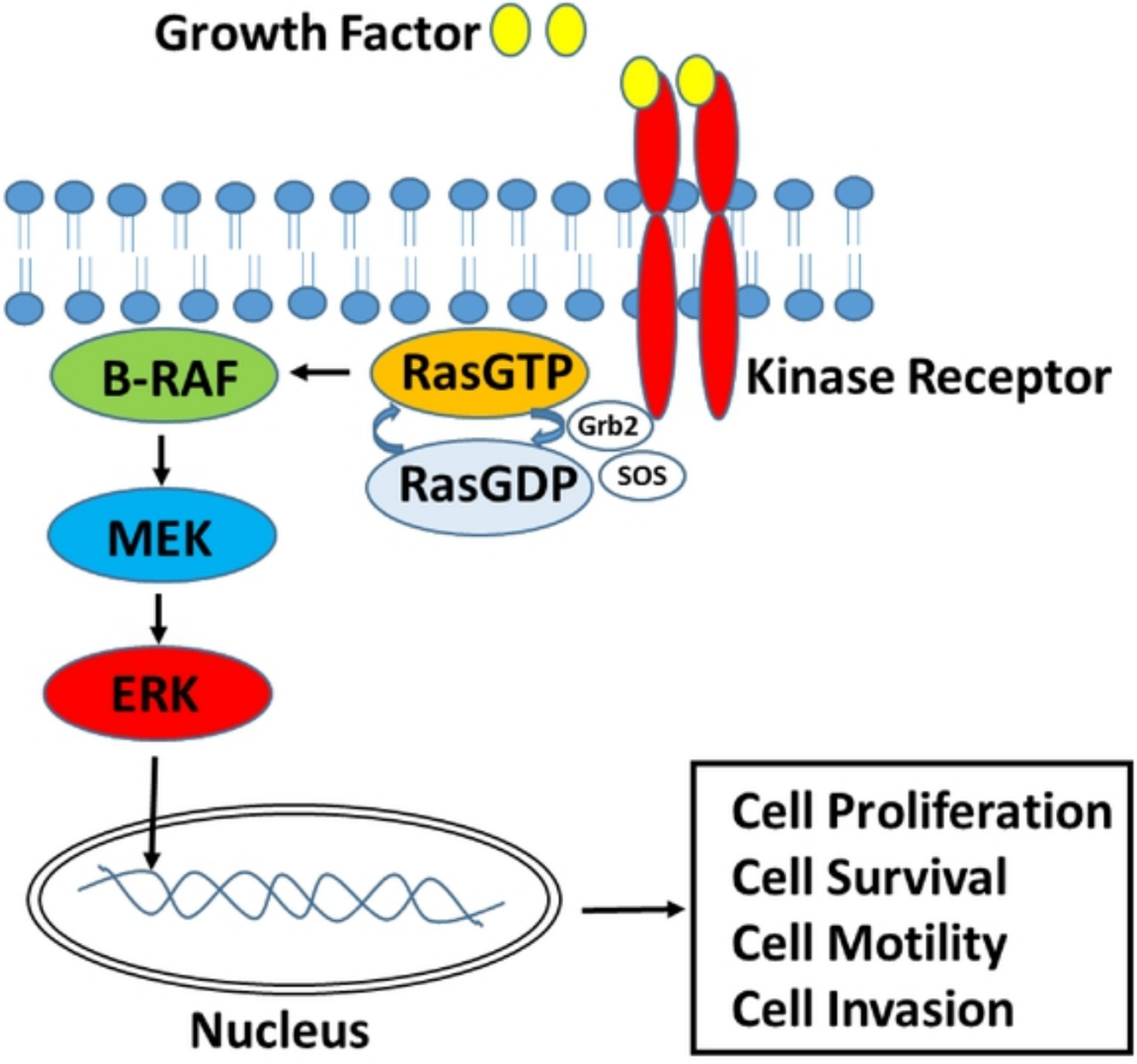

Growth Factor
Kinase Receptor
B-RAF
RasGTP
Grb2
RasGDP
SOS
MEK
ERK
Nucleus
Cell Proliferation
Cell Survival
Cell Motility
Cell Invasion

Fig 1

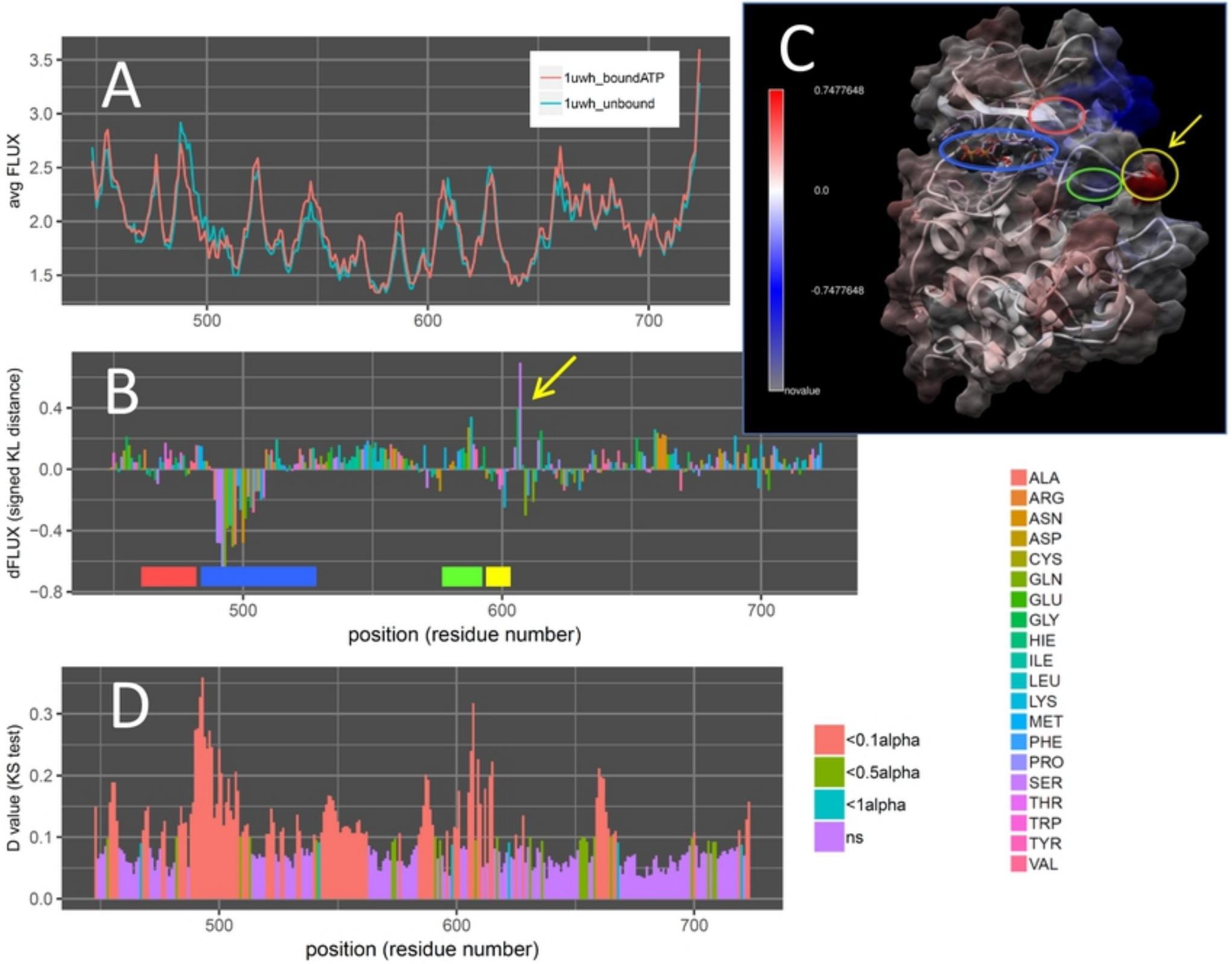
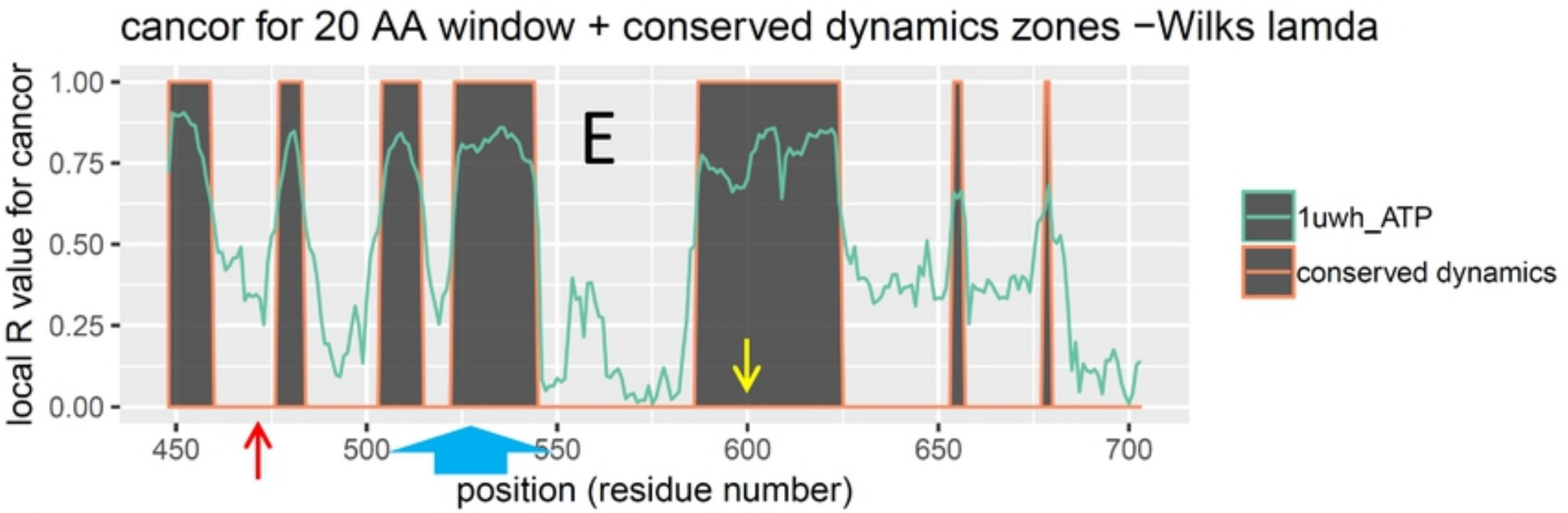
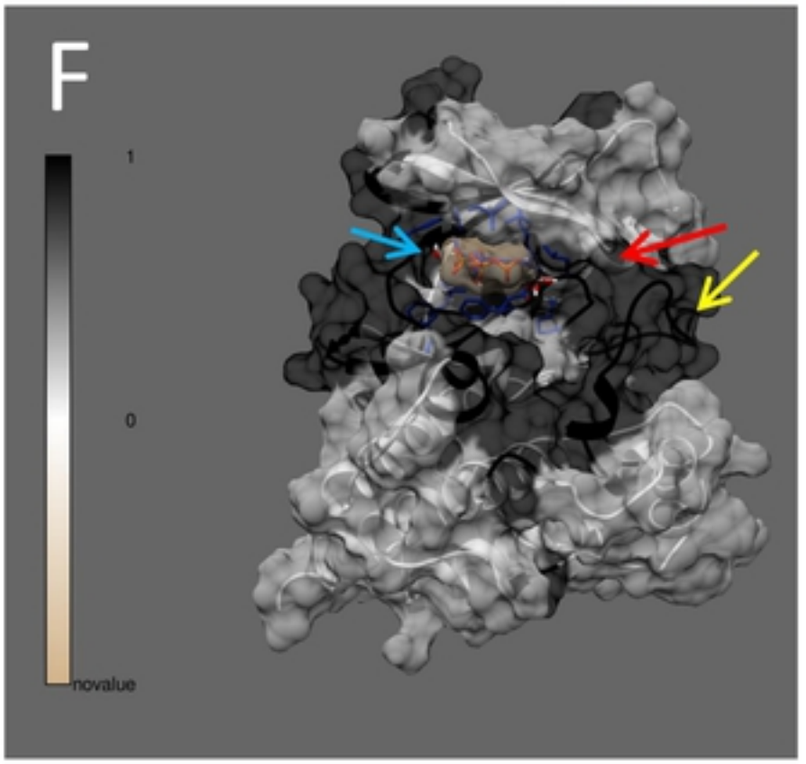

A
1uwh_boundATP
1uwh_unbound
avg FLUX
3.5
3.0
2.5
2.0
1.5
500
600
700
B
dFLUX (signed KL distance)
0.4
0.0
−0.4
−0.8
500
600
700
position (residue number)
C
0.7477648
0.0
-0.7477648
novalue
D
D value (KS test)
0.3
0.2
0.1
0.0
500
600
700
position (residue number)
<0.1alpha
<0.5alpha
<1alpha
ns
ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIE
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL

Fig 2 - top


cancor for 20 AA window + conserved dynamics zones −Wilks lamda
E
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
1uwh_ATP
conserved dynamics


F
novalue

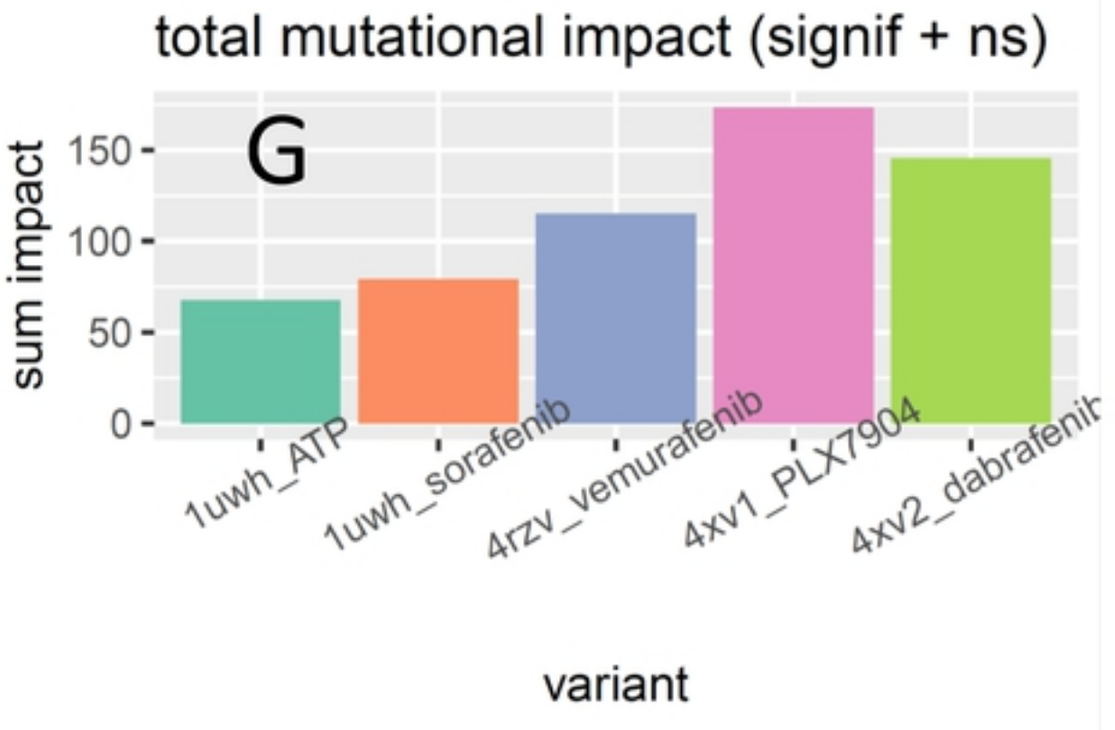

total mutational impact (signif + ns)
G
sum impact
150
100
50
0
1uwh_ATP
1uwh_sorafenib
4rzv_vemurafenib
4xv1_PLX7904
4xv2_dabrafenib
variant

Fig 2 - bottom

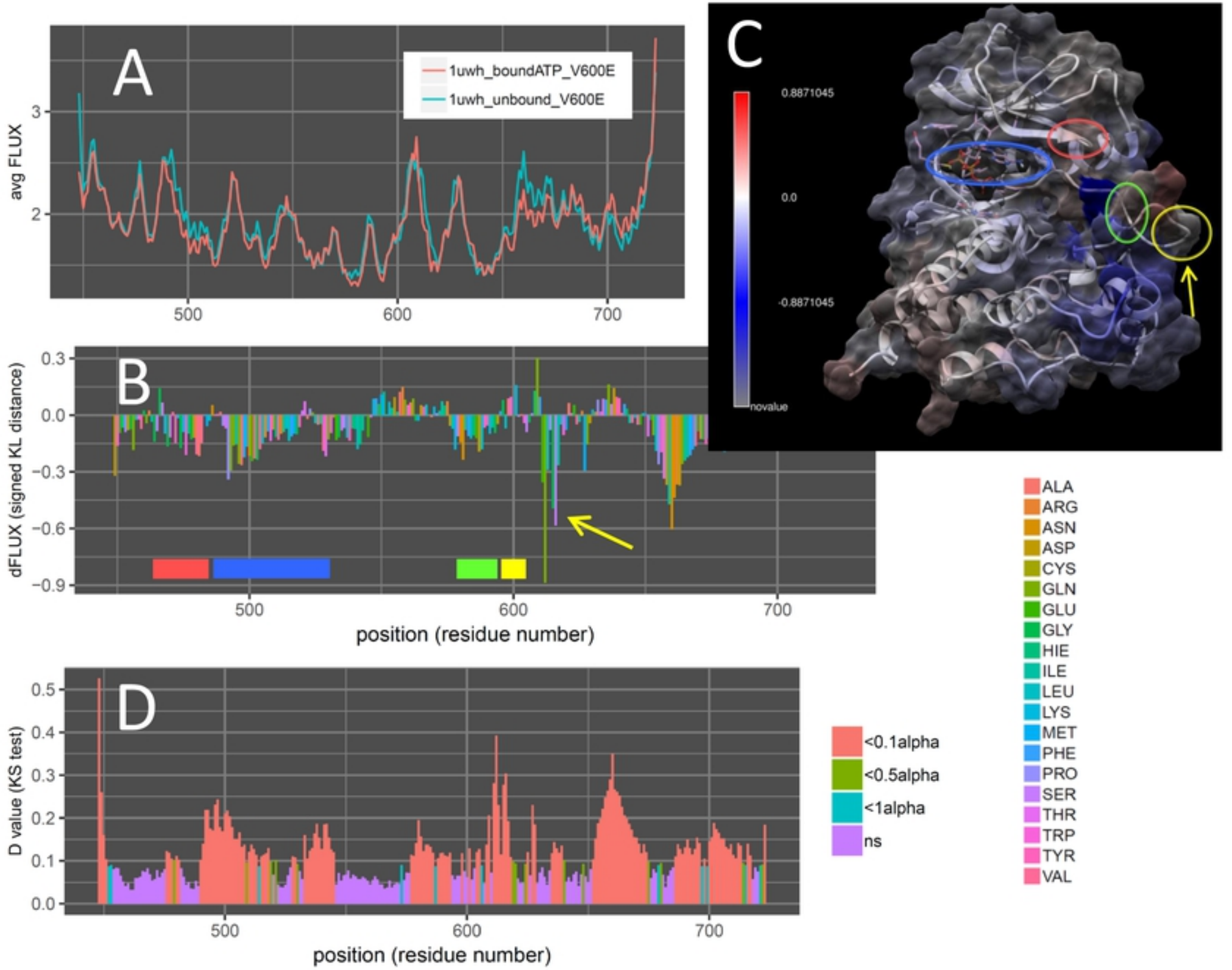
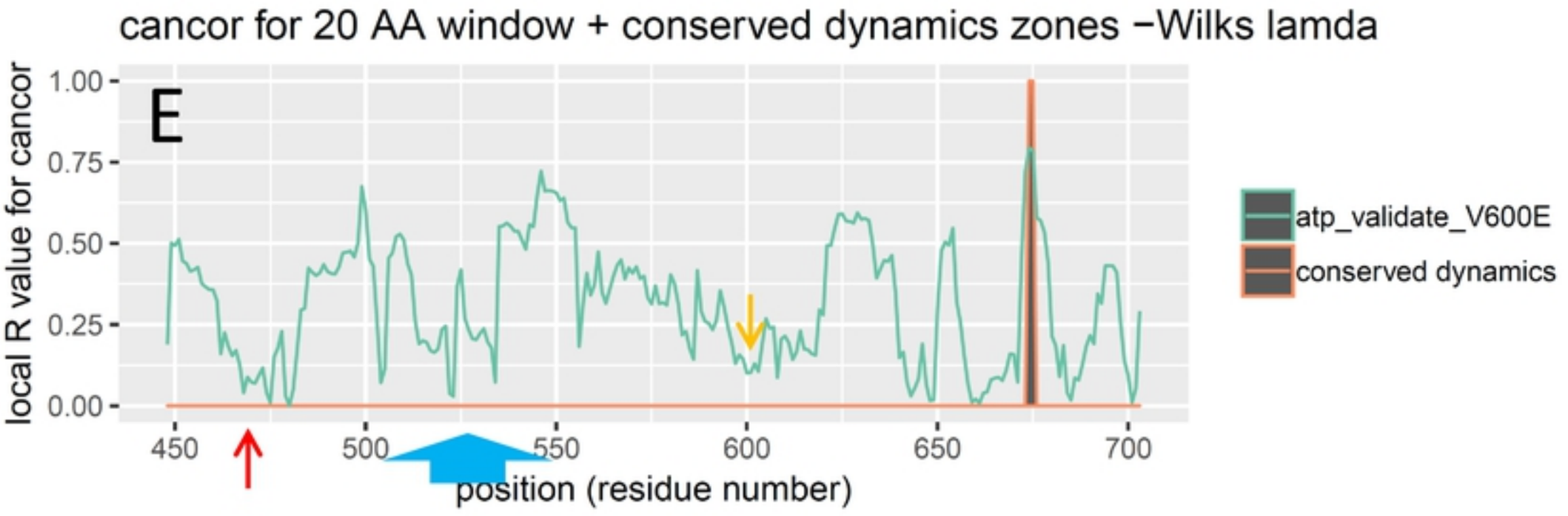
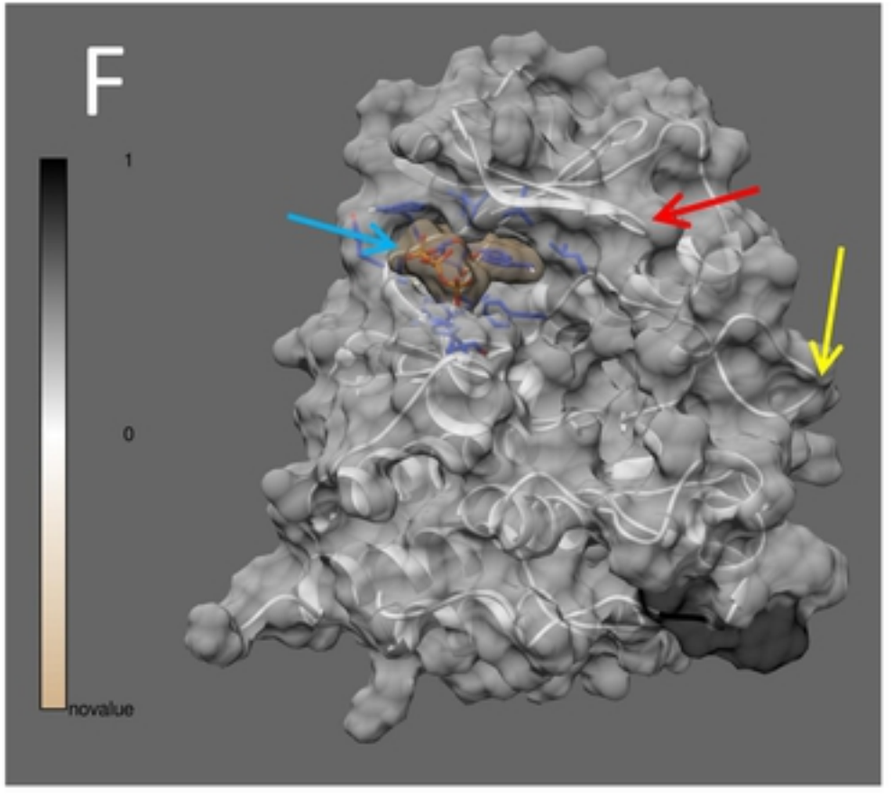

A
1uwh_boundATP_V600E
1uwh_unbound_V600E
avg FLUX
3
2
500
600
700
B
dFLUX (signed KL distance)
0.3
0.0
−0.3
−0.6
−0.9
500
600
700
position (residue number)
C
0.8871045
0.0
-0.8871045
novalue
D
D value (KS test)
0.5
0.4
0.3
0.2
0.1
0.0
500
600
700
position (residue number)
<0.1alpha
<0.5alpha
<1alpha
ns
ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIE
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL

Fig 3 - top


cancor for 20 AA window + conserved dynamics zones −Wilks lamda
E
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
atp_validate_V600E
conserved dynamics


F
1
0
novalue

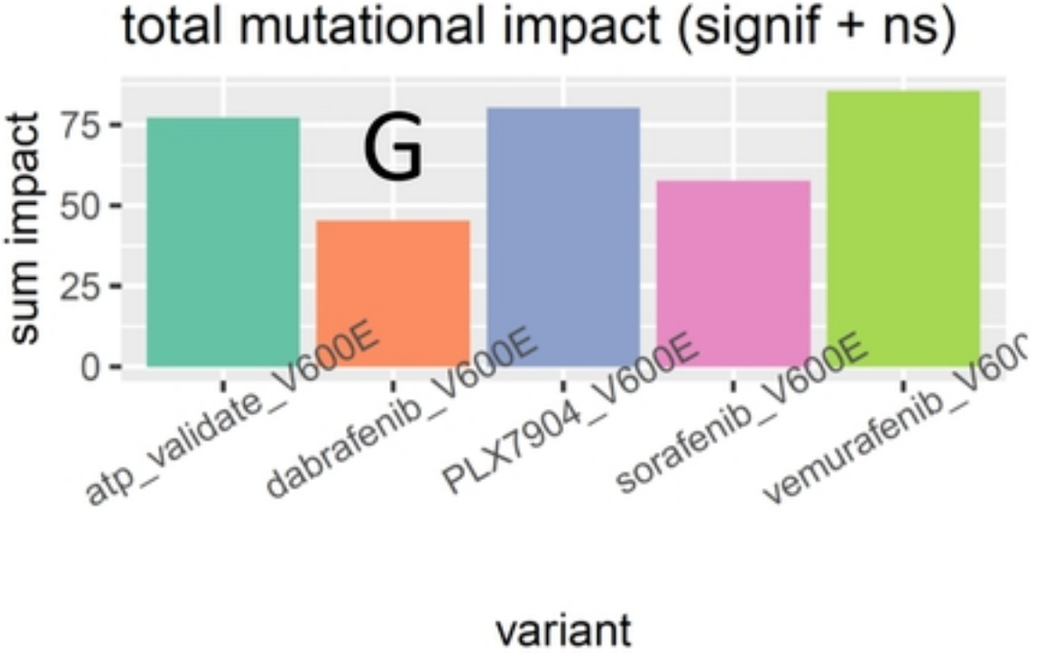

total mutational impact (signif + ns)
G
sum impact
75
50
25
0
atp_validate_V600E
dabrafenib_V600E
PLX7904_V600E
sorafenib_V600E
vemurafenib_V60
variant

Fig 3 - bottom

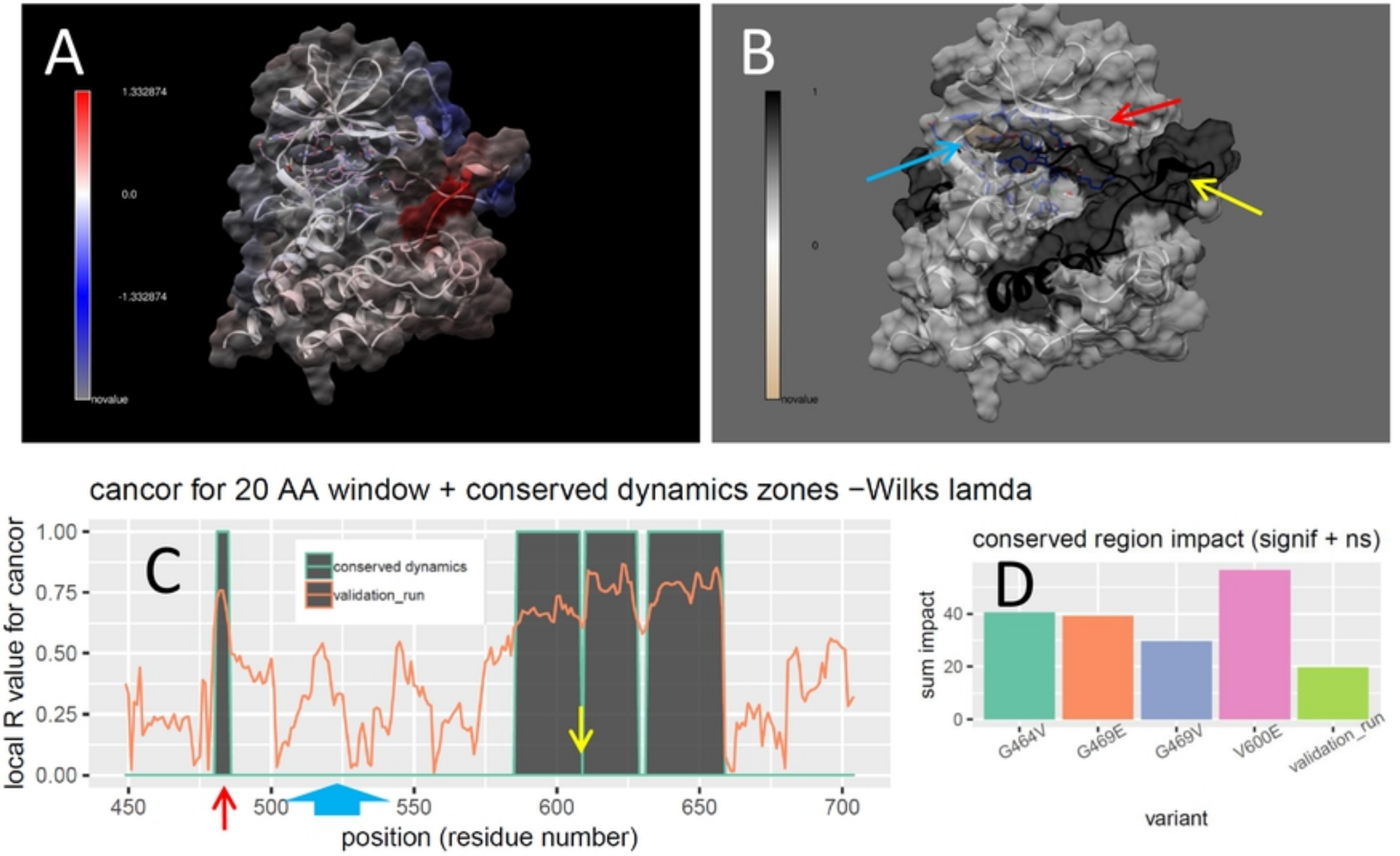

A
1.332874
0.0
-1.332874
novalue
B
1
0
novalue
cancor for 20 AA window + conserved dynamics zones −Wilks lamda
C
conserved dynamics
validation_run
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
conserved region impact (signif + ns)
D
sum impact
40
20
0
G464V
G469E
G469V
V600E
validation_run
variant

Fig 4

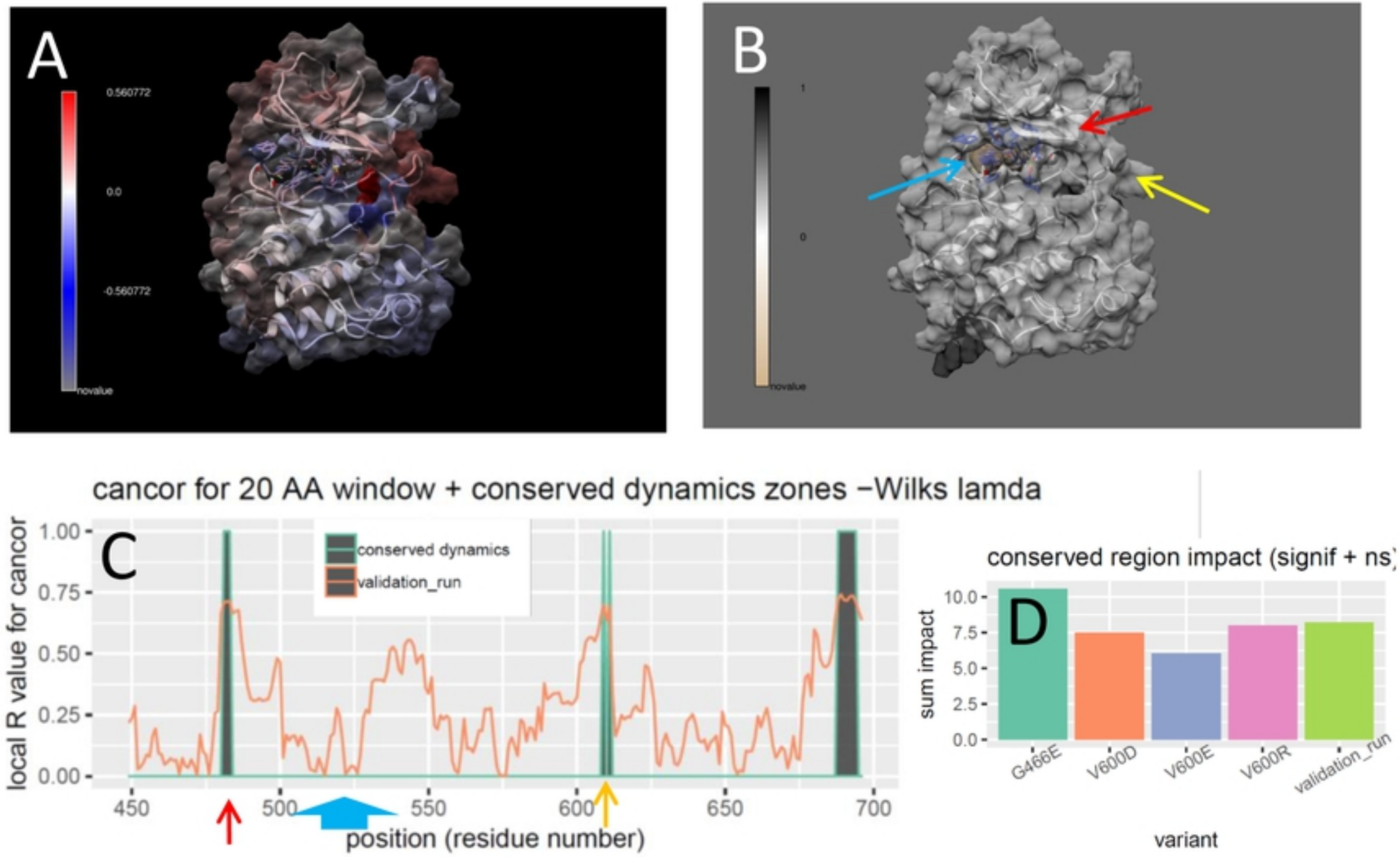

A
0.560772
0.0
-0.560772
novalue
B
1
0
novalue
cancor for 20 AA window + conserved dynamics zones −Wilks lamda
C
conserved dynamics
validation_run
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
conserved region impact (signif + ns)
D
sum impact
10.0
7.5
5.0
2.5
0.0
G466E
V600D
V600E
V600R
validation_run
variant

Fig 5

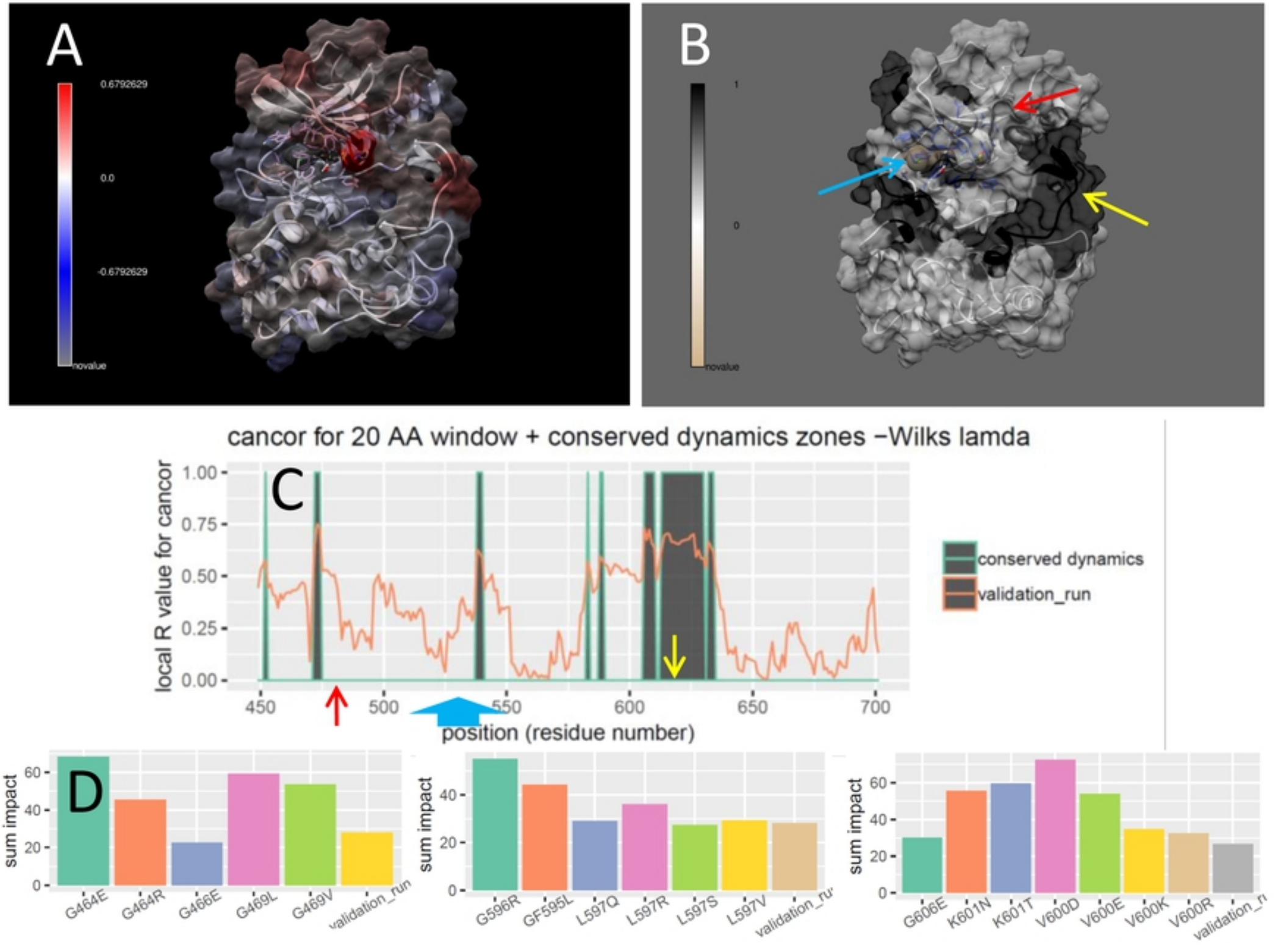

A
0.6792629
0.0
-0.6792629
novalue
B
1
0
novalue
cancor for 20 AA window + conserved dynamics zones −Wilks lamda
C
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
conserved dynamics
validation_run
D
sum impact
60
40
20
0
G464E
G464R
G466E
G469L
G469V
validation_run
G596R
GF595L
L597Q
L597R
L597S
L597V
validation_ru
G606E
K601N
K601T
V600D
V600E
V600K
V600R
validation_ru

Fig 6

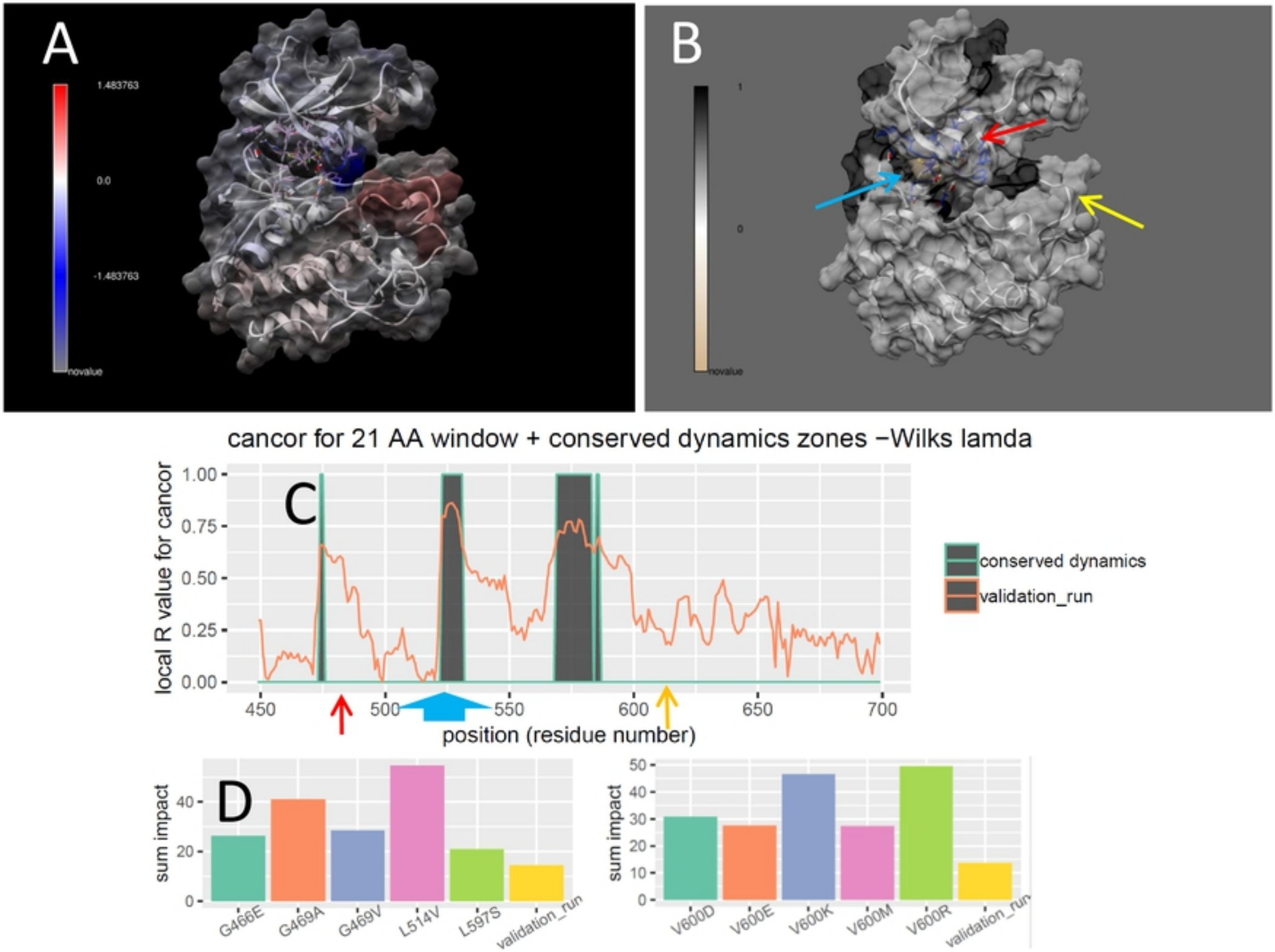

A
1.483763
0.0
-1.483763
novalue
B
1
0
novalue
cancor for 21 AA window + conserved dynamics zones −Wilks lamda
C
local R value for cancor
1.00
0.75
0.50
0.25
0.00
450
500
550
600
650
700
position (residue number)
conserved dynamics
validation_run
D
sum impact
40
20
0
G466E
G469A
G469V
L514V
L597S
validation_run
sum impact
50
40
30
20
10
0
V600D
V600E
V600K
V600M
V600R
validation_run

Fig 7

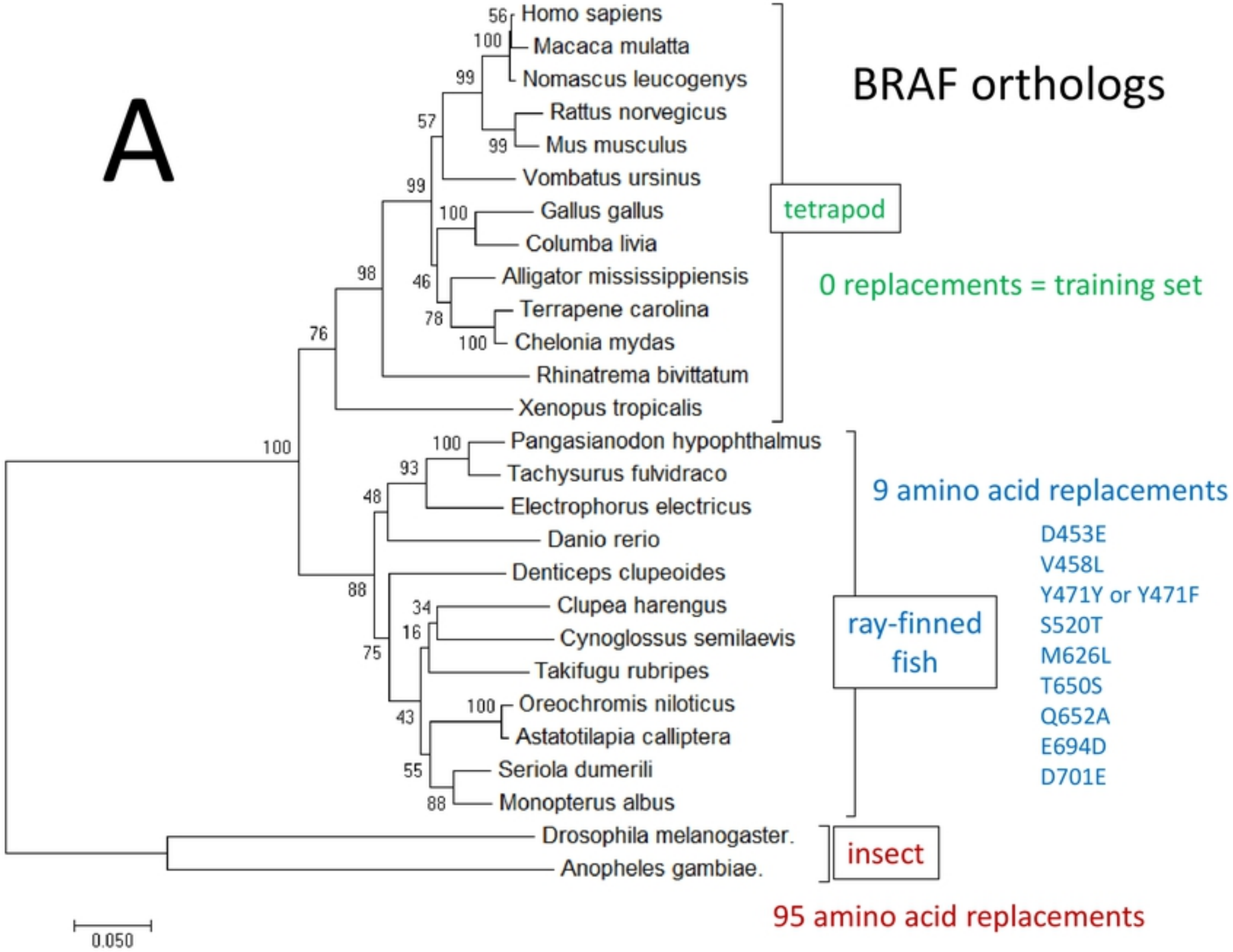
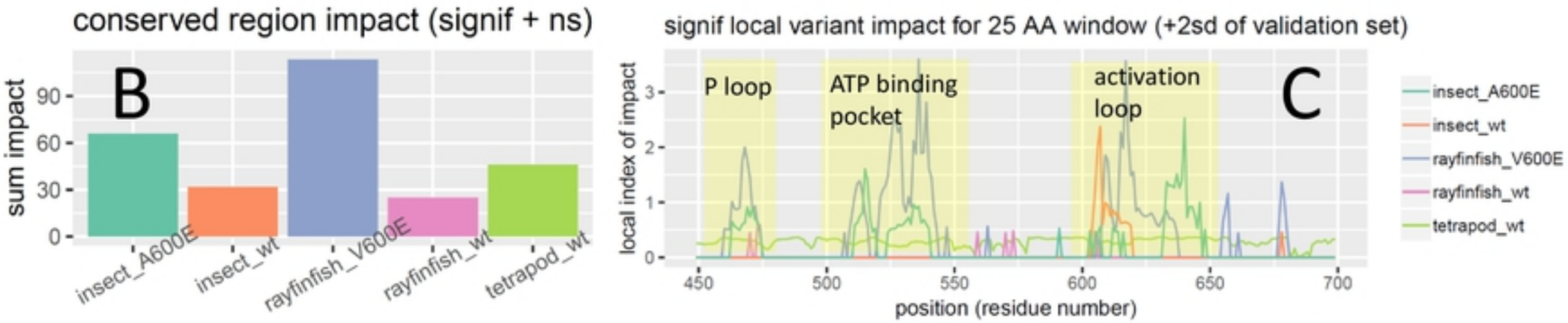

A
BRAF orthologs
Homo sapiens
Macaca mulatta
Nomascus leucogenys
Rattus norvegicus
Mus musculus
Vombatus ursinus
Gallus gallus
Columba livia
Alligator mississippiensis
Terrapene carolina
Chelonia mydas
Rhinatrema bivittatum
Xenopus tropicalis
tetrapod
0 replacements = training set
Pangasianodon hypophthalmus
Tachysurus fulvidraco
Electrophorus electricus
Danio rerio
Denticeps clupeoides
Clupea harengus
Cynoglossus semilaevis
Takifugu rubripes
Oreochromis niloticus
Astatotilapia calliptera
Seriola dumerili
Monopterus albus
ray-finned fish
9 amino acid replacements
D453E
V458L
Y471Y or Y471F
S520T
M626L
T650S
Q652A
E694D
D701E
Drosophila melanogaster.
Anopheles gambiae.
insect
95 amino acid replacements
0.050

Fig 8 - top


conserved region impact (signif + ns)
B
sum impact
90
60
30
0
insect_A600E
insect_wt
rayfinfish_V600E
rayfinfish_wt
tetrapod_wt
signif local variant impact for 25 AA window (+2sd of validation set)
C
P loop
ATP binding pocket
activation loop
local index of impact
3
2
1
0
450
500
550
600
650
700
position (residue number)
insect_A600E
insect_wt
rayfinfish_V600E
rayfinfish_wt
tetrapod_wt

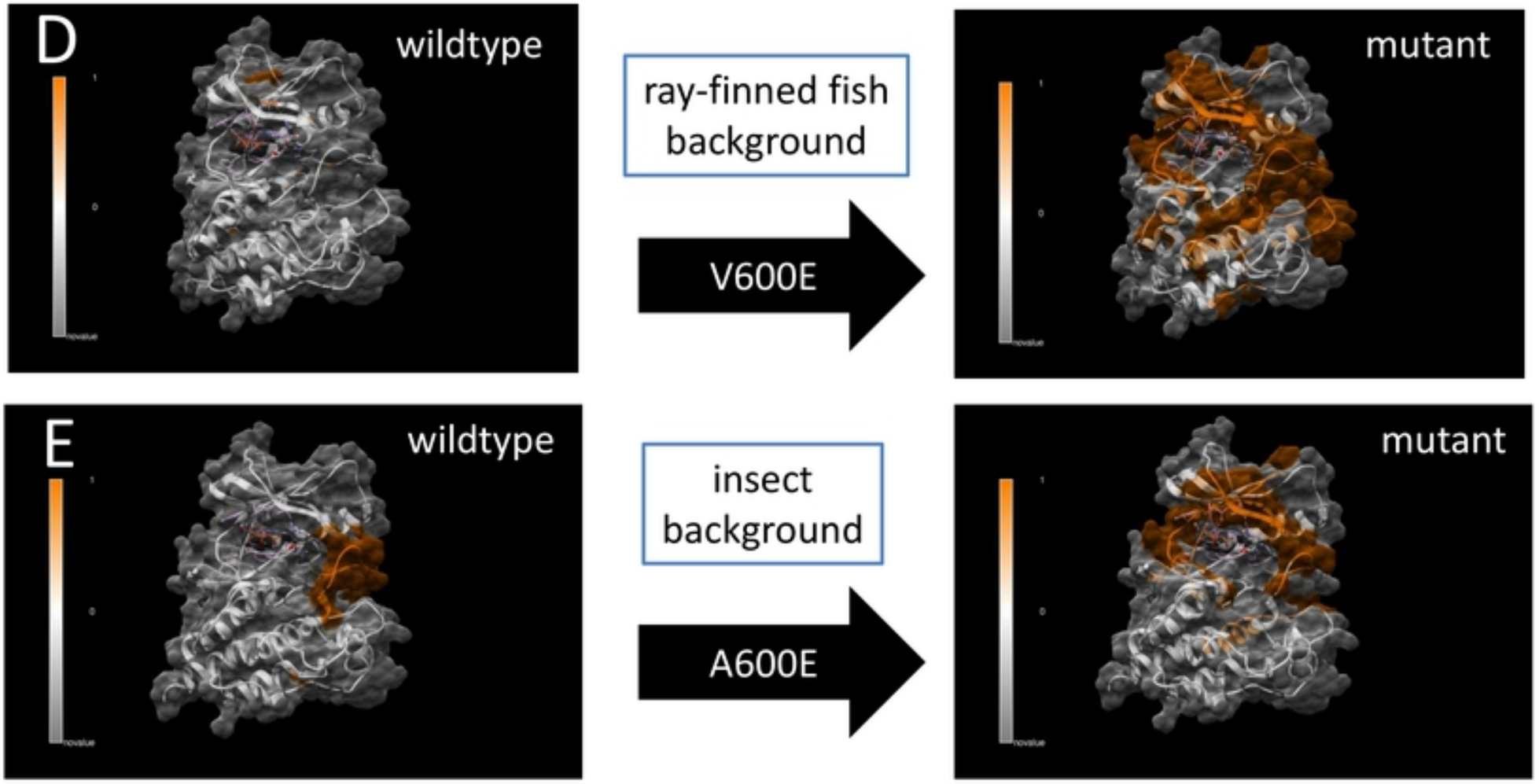

D
wildtype
ray-finned fish background
V600E
mutant
E
wildtype
insect background
A600E
mutant

Fig 8 - bottom