Deep Learning of Cellular Metabolic Flux Distributions Predicts Lifespan

Tyler A.U. Hilsabeck[1,2] and Shane L. Rea[3,*]

Affiliations:

[1]The Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, CA 94945

[2]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

[3]Department of Pathology, University of Washington, Seattle, WA 98195, USA.

*Address correspondence to: Shane L. Rea, PhD; Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA. ph. (206)685-9642; e-mail: Shanerea3@gmail.com

ABSTRACT

It is a common observation that individuals within a species age at different rates. Variation in both genetics and environmental interaction are generally thought responsible. Surprisingly, even genetically identical organisms cultured under environmentally homogeneous conditions age at different rates, implying a more fundamental cause of aging. Here we have examined the basis for lifespan variance in haploid, single-celled yeast of *Saccharomyces cerevisiae*. The probabilistic nature of metabolism means metabolites often, but not always, follow the same route through the metabolic network. We speculate redundancy in metabolic pathway choice is sufficient to explain lifespan variance. To interrogate the reaction flux space of *S. cerevisiae* we used a model of its intermediary metabolism, comprising 1,150 genes, 4,058 reactions, and 2,742 metabolites (yeast GEM_v8.5.0). We restricted traffic through the metabolic network by knocking out each of the 1,150 genes, then generated a total of 406,500 flux distributions spanning the solution space of the resulting 812 viable mutants. We collected replicative life span (RLS) data for the 812 viable mutants, corresponding to 66,400 individual cells. Four approaches were then employed to test whether reaction flux configuration could be used to predict lifespan: Principal Component Analysis (PCA) in conjunction with non-linear modeling of RLS; deep learning of RLS using either a Regression Neural Network (RNN) or a Classification Neural Network (CfNN); and deep learning using a convolutional neural network (CNN) following conversion of flux distributions to pixelated images. The four approaches reveal a core network of highly correlated reactions controlling aging rate that is sufficient to explain all lifespan variance. It includes biosynthetic pathways encompassing ceramides, monolysocardiolipins, phosphoinositides, porphyrin and glycerolipids. Our data lead to two novel conclusions. First, variance in the replicative lifespan of *S. cerevisiae* is an emergent property of its metabolic network. Second, there is convergence among metabolic configurations toward three meta-stable flux states – one associated with extended life, another with shortened life, and a third with wild type life span.

One Sentence Summary: Traffic routes and rates through the metabolic network of *S. cerevisiae* fully account for variance in replicative lifespan.

INTRODUCTION

Age is a primary risk factor for debilitating diseases such as Alzheimer's Disease, Parkinson's Disease, cancer, cardiovascular dysfunction, and osteoporosis[1–5]. In recent years, significant effort has identified conserved genetic- and biochemical processes that modulate the rate of aging in a variety of organisms. Despite such progress, our understanding of aging and how it impacts lifespan and disease incidence remains far from complete. In broad strokes we know that aging occurs at different levels in an organism: At a subcellular level, we know, for example, that key quality control processes such as protein turnover and DNA repair become increasingly compromised with time. At a cellular level, terminally differentiated cells such as neurons are lost without replacement. And, at a systemic level, we know that some tissues are likely to fail before others - the gut and vasculature, for example, are age-dependent weak points in the environmental barrier of humans. As one organ begins to fail, stereotypical compensatory responses are activated, but one damning sequelae of aging is that compensatory robustness narrows with time (homeostenosis).

Less well appreciated in the process of aging is the role of stochasticity. This is best illustrated by considering a population of genetically identical individuals cultured under environmentally homogeneous conditions – not all individuals die at the same time. Indeed, survival curves for such organisms are usually just scaled versions of those measured for higher organisms. Stochasticity comes in many flavors: At a quantum level, thermal fluctuations place atoms and molecules in different energetic states. In cells, thermal motion places molecules at random points in space. These probabilistic events are almost impossible for us to control but their effects are felt whenever life depends on a countable number of molecules, such as, for example, fate choice by transcription factors. Higher up the stochastic ladder, we know that unwanted chemical modifications accumulate with time, and these events almost certainly underlie some of the reductions in the overall efficiency of protein turnover and DNA repair that was alluded to earlier. Higher up still, we know that there are certain kinds of chance events that potentially greatly affect organismal aging, but which can be controlled to some extent. Flux through the reactions of intermediary metabolism is, in every sense, a probabilistic process and metabolites are confronted with great choice inside cells. Cells can lower this choice by placing weights on different reaction pathways – that is, by controlling the expression level of catalysts, by controlling the subcellular localization and co-assembly of such catalysts, and through metabolite shuttling where compounds are forced fed between related catalysts[6]. In this way, stochasticity can be reduced.

In this study we examined the extent to which choice of flux route through the entire network of intermediary metabolism is a significant predictor of lifespan outcome in clonal populations of the single-celled yeast, *S. cerevisiae*. Lifespan in this species can be followed by measuring the capacity of individual mother cells to generate progeny via asymmetric budding (termed replicative lifespan[7]). We find that redundancy in how compounds are fluxed through metabolic pathways is sufficient to account for all variance in *S. cerevisiae* lifespan.

RESULTS

Defining the Metabolic Solution Space of S. cerevisiae

The metabolism of *S. cerevisiae* can be represented by a multi-dimensional space (*n-space*), with flux through each reaction denoted as a dimension. Points in this space correspond to metabolic configurations, and those specific points that lead to no net change in intracellular metabolite concentrations while metabolites flux through the metabolic network correspond to positions of steady-state equilibrium. Due to physicochemical and biological restraints, the flux potential of the metabolic network is a bounded space. To determine whether there is a relationship between position in metabolic *n-space* and lifespan (**Fig. 1a**), we utilized the yeast consensus genome-scale model (GEM)[8], version 8.5.0[9], which combines 4,058 biochemical reactions across 14 sub-cellular compartments, 2,742 metabolites, and 1,150 genes, to provide an *in silico* description of *S. cerevisiae* (*strain S288c*) metabolism[8]. The 1,150 genes encode enzyme catalysts and metabolite transporters and represent ~1/5 of the yeast genome (https://www.uniprot.org). Knockout mutants of metabolic genes naturally limit the metabolic configurational space of wild type yeast. Such mutants provide a tool by which different regions of wild type configurational space can be probed to determine their relationship with lifespan. To interrogate how removing genes from *S. cerevisiae* restricts its network configurational space, we used constraint-based modeling[10]. We limited our analysis of Yeast GEM v8.5.0 to aerobic growth on glucose under defined minimal media conditions and used flux balance analysis (FBA) to identify network-wide reaction fluxes that placed the metabolic network into a steady state equilibrium when maximizing biomass production, a proxy for yeast growth. Using this approach we identified 812 viable yeast knock-out mutants that had non-zero biomass production values. Since the stoichiometric matrix for the metabolic network in *S. cerevisiae* is underdetermined, flux variability analysis (FVA) was used to chart the boundaries of the FBA solution space for each viable mutant (**Fig. 1b)**. The resulting FVA-constrained flux space (**Fig. 1c**), although bounded, effectively contains an infinite number of solutions. We therefore used Monte Carlo sampling (**Fig. 1d**) to identify a set of configurations for each knockout mutant that evenly spanned their flux space. We then removed reactions across all configurations that had either zero variance across all mutants or had variances less than 10^{-10} flux units (**Fig. 1e, f**). This resulted in a total of 406,500 flux configurations among a reduced set of 2,912 reactions that spanned the combined solution space of all 812 viable single-gene mutants.

Replicative Life Span

Replicative lifespan (RLS) analyses were undertaken manually using microdissection for all 812 viable knockout mutants. A total of 66,393 individual cells were followed until replicative exhaustion, with approximately equal numbers of mutant and pair-matched wild type cells measured during every assay. The number of individual cells for which RLS measurements were recorded varied by knockout and ranged from five to 1,717 (**Fig. 1g**). The 36,000 RLS measurements for wild type cells permitted accurate mapping of its RLS outlier boundaries (**Fig. 1h**), meaning knockout mutants with RLS *averages* falling beyond these values were low probability chance events (*p-value* <0.0035), even for knockout mutants with low RLS counts. If metabolic flux distribution is predictive of RLS, then we anticipate the RLS of any knockout mutant should be no larger or no smaller than the most extreme wild type RLS values, since the metabolism of each mutant is a subset of wild type. Indeed, RLS values of all knockout yeast fell within the span of wild type yeast RLS values (**Fig 1i,** normalized RLS values (ΔRLS) are

shown**)**. Since we were unable to assign ΔRLS values to individual configurations within the flux space of each mutant, we associated each mutant's average ΔRLS value to all its configurations. Mutual information calculations showed that all 2,912 reactions with variance across the 406,500 flux distributions held lifespan-related information, but this information content was not distributed evenly (**Fig. 1j**). In subsequent sections we extracted this information to predict yeast survival. A summary of the metabolic and RLS data employed for this study is provided in **Fig. 1k**.

Principal Component Analysis and Non-Linear Modeling of ΔRLS

To determine if position in metabolic configuration space can predict ΔRLS, we first used Principal Component Analysis (PCA) in conjunction with a Generalized Additive Model (GAM). PCA was used to further reduce the dimensionality of configurational space - from 2,912 reactions to 526 Principal Components (PCs), where each PC is comprised of a weighted contribution (loading score) of each of the original reactions. 99% of the data variance was captured by these 526 PCs (**Fig. 2a**). We next utilized a Generalized Additive Model to fit ΔRLS values using all 526 PCs, but this approach failed for three reasons – first, we had insufficient computational power to test the full model (812 ΔRLS values, 406,500 data points and 526 PCs); second, the frequency distribution of ΔRLS values was not even (**Fig. 2b**), meaning the GAM focused heavily on fitting the most frequent data; and finally, assignment of a mean ΔRLS value to all 500 configurations of each mutant confounded the spline fitting procedure because configurations that overlapped in n-space were assigned more than one ΔRLS value. We therefore refined our PCA/GAM modeling approach, first by testing whether reactions with only the highest flux rates (high flux backbone, refer to **Fig. 1e**) held ΔRLS-differentiating information (they did not), and then by modeling only the centroid configuration of each mutant. Also, we binned centroids to weight the data more evenly before GAM fitting (**Fig. 2c**). Finally, to overcome the computational limitation, we established an algorithm (*GAMRandomStep)* that allowed us to both identify PCs with predictive power (while only screening a fraction of the PCs at any one time) and to find the optimal number of PCs that maximized the fitting of each model (**Fig. 2d**). In the absence of binning, the best centroid-fitting model identified by *GAMRandomStep* was comprised of 8 PCs and used a search size of twenty-five PCs per iteration (hereafter referred to as PCA_cen). 10,000 randomizations of the ΔRLS values among the rotated centroid configurations showed the amount of deviance explained by this model (18.8%) was significant at 9.34σ (**Fig. 2e**). Although this model contained ΔRLS-predicting information (**Fig. 2f**) it performed less favorably when attempting to predict ΔRLS for all 406,500 configurations (**Fig. 2g**). Binning of the centroids led to increasingly better fitting models (**Fig. 2c**); with 100 evenly spaced bins (resulting in 71 occupied bins, **Fig. 2h**) yielding the two best fitting models (**Fig. 2c, i, j**). The amount of deviance explained by the best fitting GAM was 83.8% (using 6 PCs). This model's predictive value partially transferred when tested on all 812 centroids (**Fig. 2k**). The splines generated by the GAM for this model, and their partial predictions of ΔRLS, are shown in **Fig. 2l** and **Fig. 2m**. Finally, the loading scores for each reaction in the *second-best* fitting GAM (71 bins, 5 PCs, 79.0% explained deviance) are presented in **Fig. 2n**. We have presented data for this model (hereafter referred to as PCA_bin), and not the best fitting model, to illustrate a key point: both models utilize different, albeit overlapping, sets of PCs, even though they have the same approximate predictive power. Approximately 20% (600) of the reactions on each PC have non-zero loading scores and so

presumably the most important reactions controlling ΔRLS are still equally covered by both sets of PCs. To further refine the reactions that control ΔRLS we noted that PC loading scores often fell into three classes, distinguishable by absolute magnitude (**Fig. 2n**, ***top row)***. We expected those reactions with the biggest difference in flux when comparing the averages of knockouts representing the top and bottom 5% of ΔRLS values to also have the greatest absolute loading score values, but this was not the case (**Fig. 2n**, ***middle row***). We therefore weighted the magnitudes of the loading score of each reaction in each PC by the average difference in reaction flux rate between mutants with the top versus bottom 5% of ΔRLS values, and subsequently applied a heuristically-determined cutoff to extract a set of 593 unique reactions (**Fig. 2n**, ***bottom row***). We presume that some or all these reactions drive changes in ΔRLS.

Modeling of ΔRLS Using Neural Networks

Neural networks provide an alternate approach for non-linear data modeling. An advantage of these algorithms is that input variables remain independent, meaning their importance to the accuracy of the final trained model can be determined by simply excluding them. Three classes of neural networks were employed to identify reactions controlling the rate of aging in *S. cerevisiae*. First, we attained limited success using regression neural networks (RNN). Specifically, after testing several thousand models in which we systematically varied the network architecture (neuron size and connectivity) and multiple hyperparameters (activation function; regularization method including neuron dropout and k-fold cross validation; batch size; learning rate; number of epochs; number of input neurons (reactions); and a custom weighting parameter that forced the modeling to focus on specific subsets of data), we concluded that an RNN could not accurately model all 406,000 mutant flux configurations. It is possible that the same limitations that negatively impacted the PCA/GAM modeling were responsible for poor RNN modeling performance, namely the frequency distribution of the ΔRLS values and our assignment of a single ΔRLS value to all 500 configurations for each mutant. However, we succeeded in using an RNN to accurately predict ΔRLS using just the centroid configurations of all 812 knockouts. Two models were particularly effective - the first employed a stacking approach where three separate RNN models were generated to fit centroid flux configurations across neighboring ΔRLS ranges. This model employed the 526 reactions with the highest-ranking mutual information scores as input variables and fit the data with an RMSE of 0.574. This model could also successfully predict individual as well as binned centroid values (**Fig. 3a**). The second RNN model used the top 536 reactions with the highest-ranking mutual information scores, 10-fold cross-validation, but no stacking. This model had average R^2 and RMSE values of 0.99998 and 9.43E-07, respectively (**Fig. 3b,** *green box*). By individually excluding each reaction from this RNN, we identified 160 reactions responsible for segregating the centroid flux configurations among the 812 knockout mutants (**Fig. 3b**, *green box, bottom panel*).

Modeling of ΔRLS Using Classification Neuronal Networks

Our inability to build an RNN that could reliably distinguish all 406,000 configurations of the 812 viable knockout mutants might reflect the limitations we just mentioned, or perhaps another factor is at play. Up to this point, we have hypothesized that every metabolic configuration is associated with its own ΔRLS value (even though in practice we were forced to assign one ΔRLS value to all 500 configurations of each mutant). This hypothesis ignores the possibility that metabolic configurational switching in cells might only occur following large flux excursions in one or more reactions. So instead, it could be that a group of configurations

‘resonate’ around a central flux configuration (that represents a local minimum of chemical potential energy, say), and that configurations belonging to each group all lead to the same ΔRLS. Under this scenario, configuration switching between groups would require substantive energy input, and groups of configurations and their associated ΔRLS values would appear quantized. This idea is analogous to multi-stable states in chemistry (**Fig 3c**)[11].

Replicative lifespan curves, like the survival curves they emulate, are inverted sigmoid in shape because the distribution of individuals that gives rise to them is bell-shaped (**Fig. 3d**). The steep walls of the bell curve could be interpreted as most individuals in a population ‘trying to be the same’. In other words, for *S. cerevisiae*, it is possible there is one major metabolic configuration that every cell in the population attempts to maintain when confronted with the same environment. The ends of the bell distribution might therefore represent individuals which have inadvertently found themselves in tangible new configurations. If it is true that groups of metabolic configurations behave as if quantized for ΔRLS, then modeling ΔRLS using them becomes a classification problem rather than one of regression.

Since we have no idea of the number of configurations that could potentially stably exist on the *S. cerevisiae* metabolic network, or indeed of the reaction flux boundaries that separate them, we explored this possibility by simply grouping the collection of 406,000 knockout configurations into ‘positive’, ‘zero’ and ‘negative’ ΔRLS clusters. We utilized a custom ΔRLS threshold grouping parameter to adjust and scan the boundaries between the three clusters, and in this way, we systematically tested several thousand classification neural networks (CfNNs) for their ability to accurately distinguish yeast knockouts and predict their replicative lifespan potential. We identified three notable sets of 5-fold cross-validated CfNN models (**Figs. 3e-g**). The first set (Set A) was trained on 197,056 flux configurations drawn predominantly from knockouts with the highest and lowest 30% ΔRLS values, and it used 526 reactions with the greatest mutual information scores as modeling variables. Set A models had an average accuracy of 87.4% (+/-0.03 SD) when tested on 49,264 unseen configurations (**Fig. 3e**). The second group of models (Set B) also used 526 reactions with the greatest mutual information scores as modeling variables, but was trained on 35,380 configurations drawn predominantly from knockouts with the highest and lowest 5% ΔRLS values. Set B models had an average test accuracy of 93.8 % (+/-0.005, SD) when tested on 8,845 unseen configurations (**Fig. 3f**). Finally, the third set of models (Set C) was like Set B, except all 2009 reactions with non-zero mutual information scores were used as modeling variables, along with some other minor differences (**Supp Table 2**, summary file). Set C models had an average test accuracy of 85.4 % (+/-0.194, SD) when tested on 8,845 unseen configurations. Shown in the middle and bottom rows of **Figs. 3e-g** are the k-fold model with the highest test accuracy. Quite remarkably, the search routine that we employed to optimize our custom ΔRLS threshold grouping parameter, and which resulted in these three, best-performing sets of CfNN models, independently converged in all three instances to define the boundaries for the ‘positive’ and ‘negative’ ΔRLS clusters as the flux configurations above the 95th percentile and below the 5th percentile, respectively (**Figs. 3e-g**, *top row*).

We increased the accuracy of our CfNN modeling predictions further by combining the individual predictions of the five models in Set B into a ‘majority vote wins’ ensemble (hereafter referred to as the ‘526_CfNN’ ensemble). We did the same also for Set C (hereafter called the

‘2009_CfNN’ ensemble). Using this approach, prediction accuracies increased to 95.0% and 97.0%, respectively. To identify reactions that were driving each ensemble’s accuracy, we individually removed the 526 or 2009 reactions used as modeling variables, and then re-determined each ensemble’s prediction accuracy. For the 526 CfNN ensemble there was a total of 115 reactions that negatively impacted accuracy, while for the 2009 CfNN ensemble there were 192 reactions (**Fig. 3h**). 78 reactions were common (**Fig. 3i**). No reaction, when removed, had an impact size on ensemble accuracy that exceeded 8%.

Modeling of ΔRLS Using Convolutional Neuronal Networks

Encouraged by the ability of CfNN classifiers to predict ΔRLS for thousands of metabolic configurations, we tested the performance of Convolutional Neural Networks (CNNs) on the same task. CNNs excel at classifier tasks requiring image recognition. We therefore converted all 43,500 flux configurations from knockouts with the highest and lowest 5% ΔRLS values into images. This was accomplished using the Image Generator for Tabulated Data (IGTD) algorithm[12]. Flux values were range scaled across the 43,500 flux configurations to 256 grey scale using double-precision, floating point numbers. The IGTD algorithm assigns reactions to pixel positions in a matrix of chosen size so that the resulting correlation coefficients between all reactions (in this case across the 43,500 flux configurations) best emulates the physical distance between all pairwise pixel combinations (**Fig. 4a**). We also established a novel metric, called the ‘Inter-InterQuartile Range (IQR)-to-Median Difference Ratio’ (IIMDR, **Fig. 4b**) that facilitated assessment of the quality of our neural net modeling. IIMDR values range from zero to one, where positive values indicate two medians being compared differ at the 5% significance level. The IIMDR metric incorporates both variance and median difference measures, such that a score of zero indicates the two medians do not differ significantly, while scores close to one indicate non-overlapping medians and minimal variance in both data sets. In this way the *certainty* of the difference between two medians is emphasized at the expense of de-emphasizing the *magnitude* of the difference. We calculated the median flux of every reaction in the 2912 reduced reaction set across all 22,500 configurations of the 45 knockouts representing the bottom 5th ΔRLS percentile, and likewise across all 21,000 configurations of the 42 knockouts representing the top 5th ΔRLS percentile. Using these values, IIMDR scores were prepared for each reaction (**Fig. 4c**).

We assessed a total of 1,260 CNN models in which we systematically altered four hyperparameters (batch size, number of filters, kernel size and regularization technique). Parameter screening differed from our earlier approaches. Instead of using k-fold cross-validation, this time 14 independent searches were undertaken, all using the same set of hyperparameter choices, but differing by the seed used to initialize the random number generator, which in turn was used to initiate the neuron weights at the beginning of the optimization routine. All models were trained on 29,000 flux configurations and tested on 14,500 unseen flux configurations. All hyperparameter choices, except two, resulted in *median* test accuracies no better than guessing (**Fig. 4d**). The best choice of hyperparameters resulted in 14 models with a median test accuracy of 76.4%. Within this group, the top five models had a *mean* test accuracy of 87.2% (+/- 0.07, SD). Notably, for almost every set of hyperparameter choices tested, at least one CNN model performed far above its relevant median reflecting the importance of the initial neuronal weights in ultimately finding useful models.

To increase the accuracy of our CNN modeling we once again adopted a 'majority vote wins' approach and established two ensembles comprising multiple models. The first was composed of five models, all established using the same set of hyperparameters (CNN_A), and the second of seven models representing the best-performing models across all 1,260 CNN models constructed and each having a test accuracy >99% (CNN_B, **Fig. 4d**, *blue* and *red arrows*, respectively). Using this approach, test accuracies increased to 96.69% and 100%, respectively. Treating each ensemble separately, and by individually removing pixels (reactions) from the starting test images, we identified reactions controlling the ability of each ensemble to correctly predict ΔRLS values using flux configurations belonging to knockouts from the top and bottom 5th percentiles. Focusing on the most powerful ensemble first, CNN_B, 395 reactions decreased prediction accuracy when individually removed (effect size > 0%). All 395 reactions are presented in log-transformed form in **Fig. 4e** (see also **Fig. 4i**, *top row*). Notably, 128 of these reactions exceeded 5% effect size, with the biggest reduction in accuracy reaching 45%. These reactions are presented without transformation in **Fig. 4f** (see also **Fig. 4i,** *second row)*. Evidently, not only are the 128 reactions differential for the two sets of knockout mutants, their position in the upper right corner of the reaction location map corresponds to the red arrows shown in the reaction correlation matrix of **Fig. 4a** and implies this group of reactions are strongly anti-correlated with another part of metabolism. We identified an almost identical set of reactions driving the prediction success of the second ensemble, CNN_A. ~~(**Supp_File_A**).~~

There are 230 reactions within the Yeast_v8.5.0 GEM that are directly impacted when all the genes encoded by the 87 knockout mutants belonging to the top and bottom 5th percentiles of ΔRLS values are removed. Given the complexity associated with neural networks[13], it is possible these 230 reactions were simply assigned their own pixel during CNN modeling, and this accounts for the striking accuracy of our ensembles. Considering first the entire set of 395 reactions that have non-zero effect sizes on CNN_B accuracy (**Fig. 4e)**, 42 of the 230 reactions impacted by the 87 KOs indeed overlap (**Fig. 4g**). This number is more than the 35 expected by chance (*hypergeometric probability, 0.008*), but far fewer than 230. Considering next only the 128 reactions with effect sizes > 5% (**Fig. 4f)**, the number of overlapping reactions is 10. This number is not different from chance expectation alone (*hypergeometric probability, 0.133*). Of these 10 reactions only one corresponds to a knockout gene that exclusively targets a single reaction which, given there are 19 such genes among the 87 total (**Fig. 4h**), this number is also not different from chance expectation (*hypergeometric probability, 0.373*).

Convergence of RNN, CfNN and CNN Models

We have identified seven models, or ensembles of models, that vary in their relative success for predicting yeast ΔRLS based on metabolic flux configuration. PCA_cen and PCA_bin used centroid configurations and a small number of principal components to predict ΔRLS. These models resulted in an explained deviance of 18.8% and 79%, respectively. The RNN_536 model used centroid configurations in conjunction with 160 discrete reactions to predict ΔRLS, and this model fit the data with an RSME of 9.43E-07. For the CfNN_526, CfNN_2009, CNN_A and CNN_B ensembles, we employed tens of thousands of individual flux configurations to model ΔRLS, either from yeast with ΔRLS values in the top and bottom 30th percentiles (CfNN), or just from the top and bottom 5th percentiles (CNN). The resulting prediction accuracies of these four models on validation data were 95%, 97%, 96.7% and 100%, respectively. Beyond the obvious differences in success rate and data subsets used for the seven

models, there are two other striking differences among the seven models in relation to the predictions they make. First, as we have already seen, the reactions driving the accuracy of the ΔRLS predictions are not always the same (compare **Fig. 3h** with **Fig. 4f**). Second, the relationship between flux configuration and ΔRLS that the models predict are of two types. Explicitly, PCA_cen, PCA_bin and RNN_536 all predict a continuous function, while the four ensembles predict a quantized function where flux configurations effectively condense to just three values: positive, zero (wild type) and negative ΔRLS. How might these apparently contradictory findings be reconciled?

The left most panel of **Fig. 5a** shows all 160 reactions responsible, to varying degrees, for accuracy of the RNN_536 model. The reactions have been plotted onto the post-IGTD reaction location matrix, which was also used to display the important reactions of the CfNN- and CNN ensembles shown in **Figs. 3h** & **4f**, respectively. Shown also in **Fig. 5a** are the Olden Importance scores for each of the 160 plotted reactions. There is a cluster of 26 reactions with strongly positive Olden Importance scores (green), meaning they largely drive success of the model. To determine if the RNN algorithm naturally grouped knockouts in its bid to locate a model that could accurately predict ΔRLS, we calculated IIMDR values for all 160 reactions using binned centroid fluxes from increasingly greater group size selections (**Fig. 5a**, five middle panels). Strikingly, the 26 reactions with strongly positive Olden Importance scores overlap perfectly with IIMDR values of highest magnitude when centroids from knockouts representing the top and bottom 5th percentile of ΔRLS values are binned. This pattern continues to hold when binning centroids from knockouts out to the top and bottom 12th percentiles of ΔRLS values and is further underscored by the analogous distribution of IIMDR values calculated for the 160 reactions when the 43,000 flux distributions from knockouts belonging to the top and bottom 5th percentile of ΔRLS values are used (**Fig. 5a**, right most panel). Evidently, RNN_536 operates largely by determining whether a centroid flux configuration belongs to one of the same three classes that the CfNN- and CNN ensembles identified (positive, zero (wild-type) and negative ΔRLS). To make its final prediction, RNN_536 must employ other nodes of either less positive or negative Olden Importance to then embellish this initial decision and turn it into a continuous-value function.

Identity of Reactions Controlling Replicative Lifespan in Yeast

Each of the seven models established in this study identified a set of reactions for use in ΔRLS prediction. The constitution of six of these sets is provided in **Supp. Table K**, along with the composition of every overlap observed between the sets, and the associated hypergeometric probability for randomly observing such an overlap. We have excluded CNN_A to reduce the complexity of the analysis. For PCA_cen and PCA_bin, the statistical analyses we provide in **Supp. Table K** uses the 1305 and 824 reactions that have loading scores residing in the top or bottom 5th percentile of at least one PC driving the respective models. There are 57 possible ways a reaction can be shared among six models. We observed reactions filling all combinations. Presented in **Fig. 5b** (*upper panel)* is a proportionate Euler diagram summarizing the distribution of shared reactions among the six models. A second Euler diagram is presented in **Fig. 5b** (*lower panel)* showing the distribution of shared reactions among the six models when a smaller reaction set for PCA_cen and PCA_bin is employed (comprised of 593 and 555 reactions, respectively). The smaller set of reactions for PCA_cen and for PCA_bin was extracted from the PCs driving each model after all loading scores were weighted using the flux difference between mutants belonging to the top and bottom 5th percentile of ΔRLS values and an empirically

determined product-score cutoff was applied (see **Fig. 2n**). It is clear from both Euler diagrams that all six models show remarkable overlap. For 55 of the 57 possible ways reactions could be shared between the six models, the number of reactions shared differed significantly from chance expectation (hypergeometric probabilities ranged from 0.0033 to zero, **Supp. Table K**). Only for overlaps between RNN_536 and PCA_cen, and RNN_536 and CfNN_2009, was the number of reactions observed (N = 54 and 22, respectively) not different from expectation.

One caveat in using hypergeometric probability calculations to assess the overlap between groups that involve greater than two populations, is the order of successive pairwise-testing influences the final probability. We therefore established a second method to determine the significance of the overlaps which we observed between the reactions identified by each model. Briefly, we simulated the starting reaction composition of the six models and then randomly withdrew combinations of reactions equal to the number of reactions required for each model's accuracy. We did this 100,000 times, and for each of the 57 possible ways to share reactions among the six models we established frequency distributions that quantified the number of reactions observed in each type of overlap. Using these distributions, we assigned a probability to the actual observed number of shared reactions. This form of significance testing addresses a question that is slightly different from the one answered using the hypergeometric probability distribution. Namely, unlike hypergeometric probability testing where the composition of one model's hits is always assigned, all reaction selections in this sampling approach are randomly chosen at the start of each iteration. Using this arguably more stringent test, we determined that 35 of the 57 possible types of overlap that we observed, still contained many more reactions than expected by chance (**Supplemental Fig. J**).

How can it be that most of the six ΔRLS models share significantly more reactions than expected by chance yet visually, as seen by assessment of the post-IGTD reaction location maps shown in **Figs. 3h, 4f** and **5a,** and from the Euler diagrams of **Fig. 5b**, there are clear distinctions in the reactions that drive each model? The exciting answer to this question is provided in **Fig. 5c**. Up to this point we have assessed the contribution of individual reactions to the predictive power of each of the six ΔRLS models using metrics specific to each modeling technique, making it hard to compare results across the separate studies. So, to do such a comparison, we established a uniform metric. First, for PCA_cen and PCA_bin we took the 593 and 555 product scores of the reduced reaction sets and range scaled them between 0 and 1. For RNN_536 we took the absolute values of all 160 nonzero Olden Importance Scores and similarly range scaled them between 0 and 1. For CfNN_526 and CfNN_2009, all non-zero 'percent reduction in ensemble accuracy' scores were converted to positive values then range-scaled between 0 and 1. And lastly, for CNN_B, all 'percent reduction in ensemble accuracy' scores > 5% were range-scaled between 0 and 1. This resulted in a total of 1,743 reactions. Focusing only on those reactions that were important to more than one model (908 in total), we assigned the range-scaled score that corresponded to the highest measured value across all relevant models and then plotted this subset of reactions onto the 54x54 reaction location matrix generated by the IGTD algorithm during CNN model construction (**Fig. 5c**, *left panel*). Most of the reactions are evidently of small impact and vanish into the background. Two broad clusters of reactions are, however, evident (*bracketed*). They are more easily discernable using a scatterplot of the range-scaled scores versus a linear ordering of the reactions (54x54 row concatenation) (**Fig. 5c**, *middle panel, asterisks)*. We selected the highest-ranking of these reactions in both clusters (range-scale

score > 0.6), then mapped the location of the resulting 90 reactions onto the post-IGTD reaction correlation matrix, which was initially derived using the 43,500 flux configurations from knockouts with ΔRLS values representing the extreme 5th percentiles (**Fig. 5c**, *right panel, arrows)*. Notably, the two clusters of reactions are strongly anti-correlated. Evidently, all six ΔRLS prediction models recognized the same changes in metabolism to predict replicative lifespan – except that either of two sets of reactions were chosen to do so because the sets are anti-correlated with each other.

To map the finer details of the relationships among the highest-ranking 90 reactions, we extracted all the correlation coefficients among these reactions from the post-IGTD correlation matrix shown in **Fig. 5c,** (*right panel)*, and then used t-distributed stochastic neighbor embedding (tSNE) to cluster the resulting 90x90 vector profiles (**Fig. 5d**). We then used these groupings to construct a heat map based on correlation coefficient magnitudes **Fig. 5e**. Notably, the two reaction clusters which we identified in **Fig. 5c** remained largely unchanged despite the obvious appearance of finer relationship structure (**Fig. 5e,** *colored text)*. To determine how the flux rate of each reaction varied with replicative lifespan, we again used the 43,500 flux configurations associated with ΔRLS values representing the extreme 5th percentiles. We constructed boxplots for each of the 90 reactions to summarize the flux differentials between knockouts with the highest and lowest ΔRLS values (**Supp Fig. L**). Reactions belonging to each tSNE group displayed similar profiles, so we selected the reaction in each group with the largest range-scaled score differential for presentation in the main figure (**Fig. 5f**, *box plots*). Focusing on the notched region in each box, (notches that do not overlap have different medians at the 5% significance level), we see that flux readings for reactions in tSNE Groups 1 and 5 are significantly higher in yeast knockouts with the highest ΔRLS values versus those with the lowest. A reversal of this flux differential is observed for reactions in tSNE Groups 2, 4, and 6. Among the reactions controlling replicative lifespan, it is noteworthy we did not identify the reaction used to represent biomass production in the Yeast 8.5.0 model (**Fig. 5g**). This implies metabolic flux configurations that are associated with altered replicative lifespan do not simply slow down or speed up growth, rather, they change the routes by which metabolites are moved through the metabolic network (see Metabolic Map, **Supplemental Data**).

Collectively, our data point toward two exciting conclusions regarding lifespan determination in yeast. First, each metabolic flux configuration appears to be associated with an inherent survival probability. In other words, length of life is an emergent property of metabolic flux configuration. Second, although there is an infinite number of flux configurations available to the yeast metabolic network (albeit constrained within physicochemical bounds), there appears to be just three pseudo-metabolic states that these configurations converge toward, and each one is associated with a major shift in survival probability (**Fig. 5h**).

DISCUSSION

As rousing as our conclusions are regarding life span determination in yeast, we must temper them and acknowledge that the Yeast 8.5.0 genome-scale metabolic model employed in this study, represents just one iteration of a community model undergoing continuing refinement. Indeed the latest GEM, Yeast 9.0, has just been released[14]. Despite its remarkable complexity, the Yeast 8.5.0 GEM contains an incomplete description of lipid metabolism (as does Yeast 9.0). Moreover, the rules that this GEM uses to determine enzyme redundancy rely, for the most part, on protein homology, but recent studies show that incorporating mRNA co-expression data leads to more accurate predictions. Lastly, for some reactions, flux bounds are knowingly broad. Newer tools that measure enzyme concentrations or estimate enzyme *kcat* values continue to refine these ranges.

In this light, the results obtained by our study should be viewed as the first step in an evolving research endeavor. Clearly, the metabolic configurations that we have predicted for each knockout mutant are approximations at best. That said, Yeast 8.5.0 and GEMs like it, continue to be used extensively in both the academic and industrial settings, and their worth in predicting changes to metabolic flux following different kinds of genetic alteration are not disputed. Just as the GEM field has made fantastic strides in iteratively refining the reconstruction of the metabolism of *S. cerevisiae*, we expect that the techniques that we have developed in this paper will serve as a foundation for using metabolism to predict lifespan in yeast and other models for which GEMs exist. As more information is incorporated into GEMs, we anticipate that the relationship between aging and metabolism will be further refined. Whether it is true that lifespan is tri-stable or n-stable, in yeast or any other organisms, remains an exciting possibility. During our study, we note that a GEM for *S. cerevisiae* that incorporate enzyme concentration (ec) constraints was introduced and, in turn, successfully used to alter heme production[15]. Also, as mentioned, Yeast 9.0 was released a couple months ago[14], and this model contains an additional 73 reactions, 62 metabolites and 13 genes, as well as newly added thermodynamic information for standard biochemical conditions. Notably, deep learning approaches have been used to predict enzyme *kcat* values to further enhance the ec-GEM model[16]. Finally, there is an end game in the GEM field to incorporate both DNA transcriptional control and RNA expression data to have a fully integrated synthetic model of cellular life (multi-scale models)[17]. Techniques that we introduce in this study also have room for improvement. For example, we are yet to devise approaches that scan for optimal assignments of ΔRLS values to individual flux configurations. For computational and cost reasons, the approach we relied upon assigned a single centroid value to all the flux configurations of each knockout.

Notwithstanding the above limitations, can we gauge the relative reliability of our current predictions? In our effort to answer this question, we were shocked to discover that 359 of the 812 single-gene knockouts used in this study (44%) had flux space assignments identical to wild

type. It is immediately clear that this observation likely explains many of the modeling challenges that we faced during our investigation. It also raises an additional flag: Our working hypothesis was survival potential is dependent upon metabolic configuration, and so we would have predicted that the replicative lifespan of these mutants would be the same as wild type, but of course we know this is not true. There is nothing extraordinary about the RLS values of the 359 KOs. We find they are scattered randomly across the full range of replicative lifespans that we measured (Supp Table: summary of actual genes). We conclude then, either our working hypothesis is wrong and there are 359 metabolic enzymes that moonlight outside of metabolism in some unknown manner to affect survival, or the flux space assignments made by the Yeast 8.5.0 GEM for these 359 KOs are inaccurate. Almost certainly the latter is true. Shortly, we shall show that although these incorrect flux space assignments did impact the final list of reactions found by our modeling approaches, the lists were augmented with a subset of random reactions that can be identified and excluded, leaving the real signal.

Specifically, the CNN_B ensemble of convolutional neural networks achieved 100% accuracy when distinguishing the 43,500 metabolic configurations belonging to knockout yeast with ΔRLS values in the highest and lowest 5th percentiles. Interrogation of the flux space predictions made by the Yeast 8.5.0 GEM for the 87 KO mutants representing this set of yeast reveal that 44 of the mutants (51%) occupy the same flux space as wild type. How then did the CNN_B ensemble achieve 100% accuracy? Analysis of Supplemental File X clarifies this answer. It appears that during training, the CNN_B ensemble learned to distinguish four groups of metabolic configurations instead of only two, as we originally believed. That is, the CNN_B network learned to segregate KOs from the two ΔRLS extremes that had real configurational differences by using reactions that truly differed between them. Simultaneously, the CNN_B network learned to distinguish KOs that were assigned wild type configurations in both groups, by using reactions that, through chance alone, had median differences that were exploitable. Recall that we used 500 configurations per KO which were randomly sampled from their FVA-defined solution space. The choice of which reactions CNN_B used to distinguish the second pair of KOs was ultimately decided at this step, specifically by the random assignment of reaction rates to each reaction. This use of four different groups becomes evident when IIMDR values are considered (refer to **Supplemental File**: SUPP_TABLE_medians_and_IIMDRs_cluster1and2_all).

Based on the process of reaction selection used by the CNN_B ensemble, we hypothesize that the reactions contained in Cluster 1 and Cluster 2 of **Fig. 5c** are likely also of two types – real and artifactual. This is because both clusters were identified by combining the results from six separate modeling approaches that all drew from the same set of 406,000 configurations. There are two approaches that theoretically could be used to differentiate both reaction types. First, if a new set of 500 configurations for each KO is generated and our entire set of modeling analyses were to be repeated, we would expect a new subset of artifactual reactions to replace the old ones, since they are ultimately randomly generated. The real reactions, however, should reappear in the new analysis. Second, if our entire analysis were to be repeated using only KOs whose flux solution space is demonstrably different from wild type, only real reactions should reappear in the new analysis. We tested the latter approach, directing our attention to just the 87 knockouts representing yeast with the highest and lowest 5% ΔRLS values. In this set of KOs, there are 43 mutants that the Yeast 8.5.0 GEM predicts have flux spaces that differ from wild

type. Since all 87 mutants have a life span phenotype, we confidently updated the redundancy rules for the subset of mutants whose gene products only affected a single reaction but were initially defined as being redundant with one or more other enzymes, but obviously were not. An additional 10 mutants were therefore identified whose flux space differed from wild type when the flux rate through the single reaction they catalyzed was set to zero (or in two cases, when reduced to a level that caused the growth rate to be 10% of wild type). Finally, 10 of the 45 mutants corresponding to yeast with ΔRLS values in the lowest 5th percentile were vacuolar membrane (H^+)-ATPase (Vma) mutants. We selected only one of these, reducing any bias that might have driven the original analysis results. Using a final collection of 44 mutants then, represented by 22,000 unique metabolic configurations, we constructed new CfNN and CNN models, isolated the reactions responsible for their accuracy, then compared the resulting reaction sets with the reactions comprising Clusters 1 and 2 (Supplemental Data). Significant overlap between the datasets was observed. Of 48 reactions that could have been identified among a set of 2473, 18 reactions in total were re-discovered (sequential hypergeometric probability, $p < 0.004$) (Supp File: non-zero CNN_2473_hits_comparisons). This limited analysis further underscores our initial conclusions, that length of life is an emergent property of metabolic flux configuration, and that there is a convergence among metabolic configurations toward meta-stable flux states. As newer GEMs become available it will be exciting to repeat our analyses in full.

Final mention should be made of the IIMDR metric that we introduced in this study. Initially established as a tool for assessing the quality of our neural net modeling, we note that this metric was also surprisingly effective at predicting which variables would ultimately be employed by our fully trained networks for object discrimination. Use of IIMDR values by the broader artificial intelligence community might therefore be useful for reducing the complexity of training data to speed up the training process.

ACKNOWLEDGEMENTS
The authors thank Matt Kaeberline, Brian Kennedy, and Mark McCormick for sharing the yeast lifespan data used in this paper. Meghan Cain and Turner Conrad performed preliminary work. T.A.U.H. was trained and supported by Pankaj Kapahi and Judith Campisi at the Buck Institute. T.A.U.H. was supported by NIH/NIA award F31AG062112, by NIH/NIA training grant T32AG000266-24, and RF1 AG 068908-02.

AUTHOR CONTRIBUTIONS
T.A.U.H. and S.L.R. designed research; T.A.U.H. and S.L.R. performed research; T.A.U.H. and S.L.R. analyzed data; T.A.U.H. and S.L.R. provided Matlab, R, and Python code; T.A.U.H. and S.L.R. wrote the paper; S.L.R. provided experimental guidance and funding.

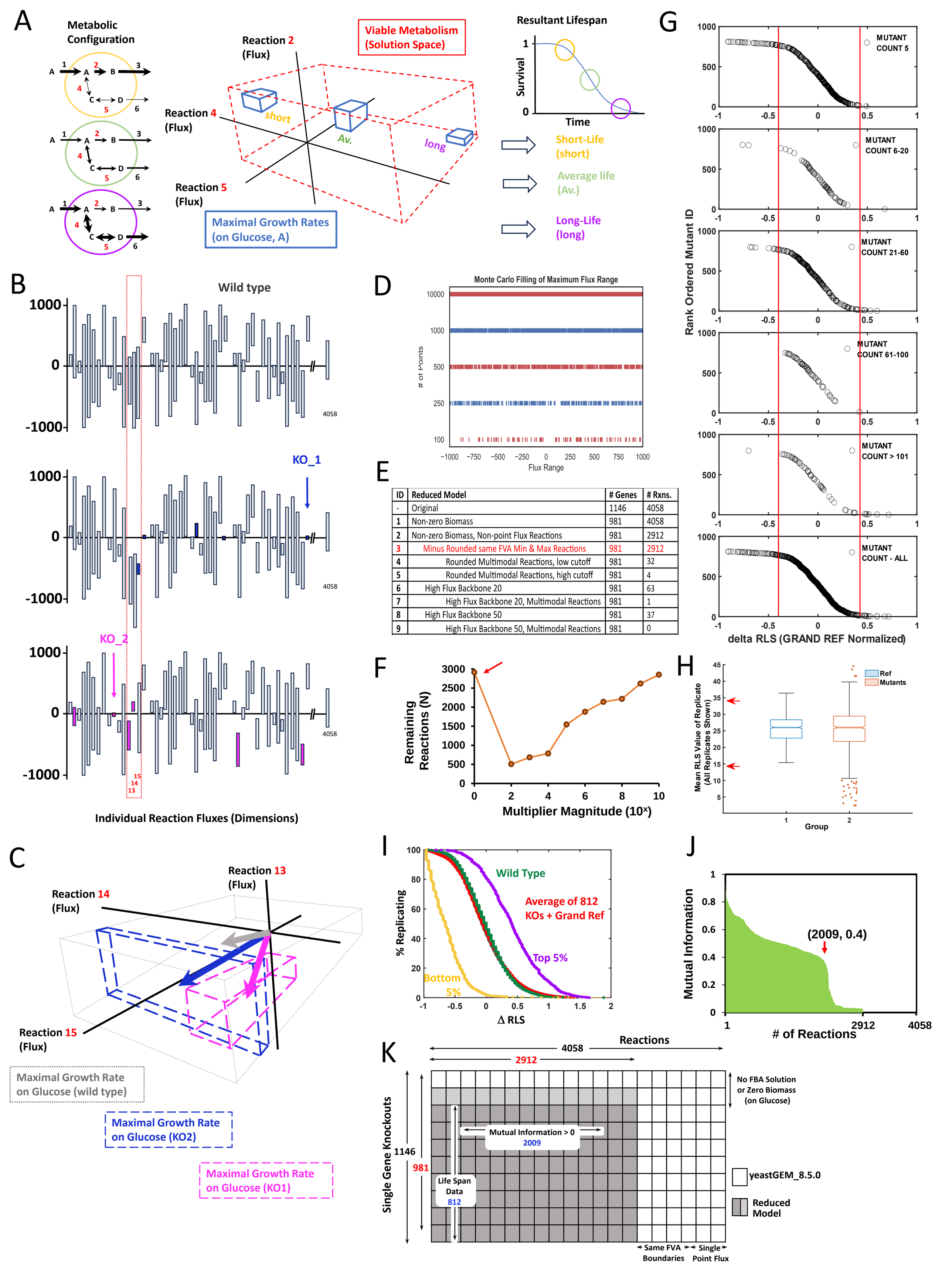
DECLARATION OF INTERESTS
The authors declare no competing interests.

DATA AVAILABILITY
The code used for this project can be found in the github repository yeast-GEM-Lifespan-Modeling (https://github.com/thilsabe/yeast-GEM-Lifespan-Modeling). Data files generated

using this code can be found at the figshare repository Hilsabeck, et. al. 2024 - Deep Learning of Cellular Metabolic Flux Distributions Predicts Lifespan (LINK).

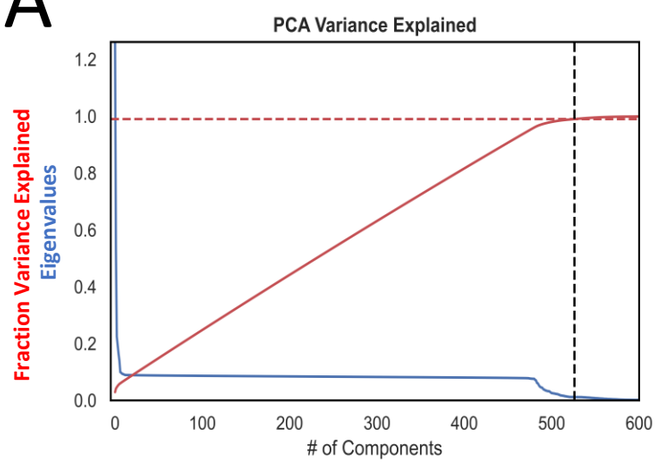
REFERENCES

1. Podolskiy, D. I., Lobanov, A. V., Kryukov, G. V. & Gladyshev, V. N. Analysis of cancer genomes reveals basic features of human aging and its role in cancer development. *Nature Communications* **7**, 12157 (2016).
2. Wang, J., Xiong, S., Xie, C., Markesbery, W. R. & Lovell, M. A. Increased oxidative damage in nuclear and mitochondrial DNA in Alzheimer's disease. *Journal of Neurochemistry* **93**, 953–962 (2005).
3. Moslehi, J., DePinho, R. A. & Sahin, E. Telomeres and Mitochondria in the Aging Heart. *Circulation Research* **110**, 1226–1237 (2012).
4. Edwards, B. K. *et al.* Annual Report to the Nation on the status of cancer, 1973–1999, featuring implications of age and aging on U.S. cancer burden. *Cancer* **94**, 2766–2792 (2002).
5. Collier, T. J., Kanaan, N. M. & Kordower, J. H. Ageing as a primary risk factor for Parkinson's disease: evidence from studies of non-human primates. *Nature Reviews Neuroscience* **12**, 359–366 (2011).
6. Cohen, Y. *et al.* Peroxisomes are juxtaposed to strategic sites on mitochondria. *Mol. BioSyst.* **10**, 1742–1748 (2014).
7. MORTIMER, R. K. & JOHNSTON, J. R. Life span of individual yeast cells. *Nature* **183**, 1751–2 (1959).
8. Lu, H. *et al.* A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* **10**, 3586 (2019).
9. Benjamín Sánchez, feiranl, E. K. eiden309, M. A. W. S. T. P. hongzhonglu, D. L. S. M. F. & T. G. Badger. SysBioChalmers/yeast-GEM: yeast 8.5.0 (v8.5.0). Preprint at (2021).
10. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* **6**, 1290–1307 (2011).
11. Ramakrishnan, N. & Bhalla, U. S. Memory Switches in Chemical Reaction Space. *PLoS Comput Biol* **4**, e1000122 (2008).
12. Zhu, Y. *et al.* Converting tabular data into images for deep learning with convolutional neural networks. *Sci Rep* **11**, 11325 (2021).
13. Hu, X., Chu, L., Pei, J., Liu, W. & Bian, J. Model complexity of deep learning: a survey. *Knowl Inf Syst* **63**, 2585–2619 (2021).
14. Zhang, C. *et al.* Yeast9: A Consensus Yeast Metabolic Model Enables Quantitative Analysis of Cellular 1 Metabolism By Incorporating Big Data 2. **637459**, 19.
15. Ishchuk, O. P. *et al.* Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* **119**, (2022).
16. Li, F. *et al.* Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* **5**, 662–672 (2022).
17. Lu, H., Kerkhoven, E. J. & Nielsen, J. Multiscale models quantifying yeast physiology: towards a whole-cell model. *Trends Biotechnol* **40**, 291–305 (2022).

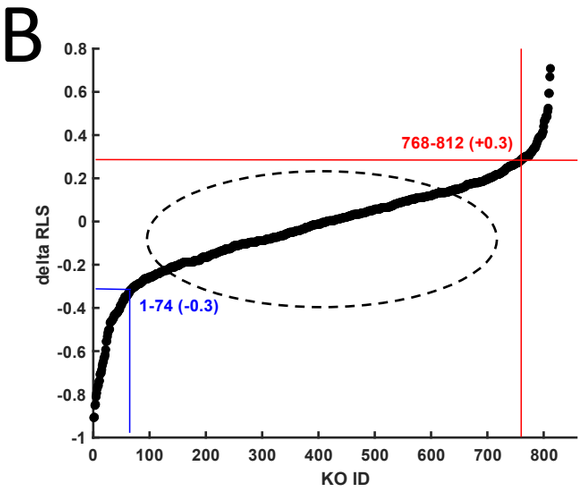


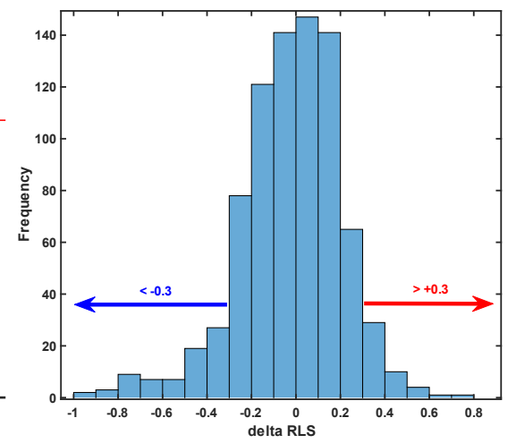
ID	Reduced Model	# Genes	# Rxns.
-	Original	1146	4058
1	Non-zero Biomass	981	4058
2	Non-zero Biomass, Non-point Flux Reactions	981	2912
3	Minus Rounded same FVA Min & Max Reactions	981	2912
4	Rounded Multimodal Reactions, low cutoff	981	32
5	Rounded Multimodal Reactions, high cutoff	981	4
6	High Flux Backbone 20	981	63
7	High Flux Backbone 20, Multimodal Reactions	981	1
8	High Flux Backbone 50	981	37
9	High Flux Backbone 50, Multimodal Reactions	981	0

A

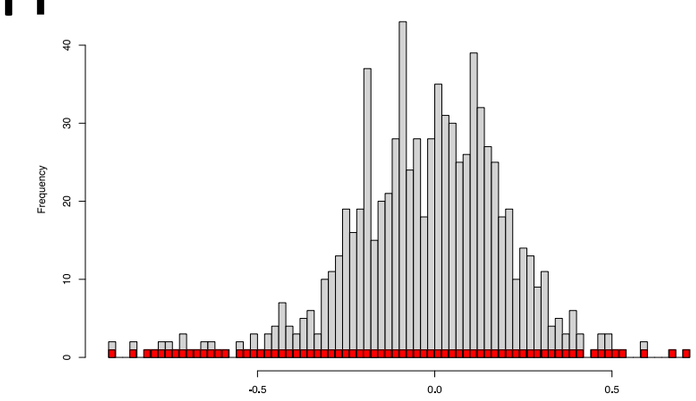


B





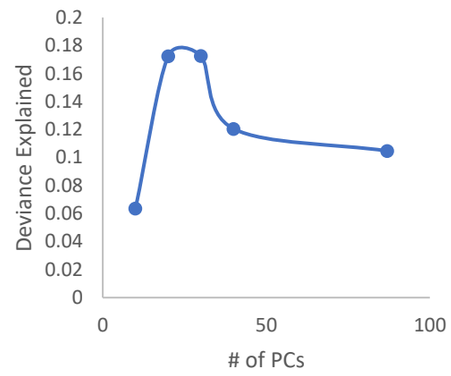
H



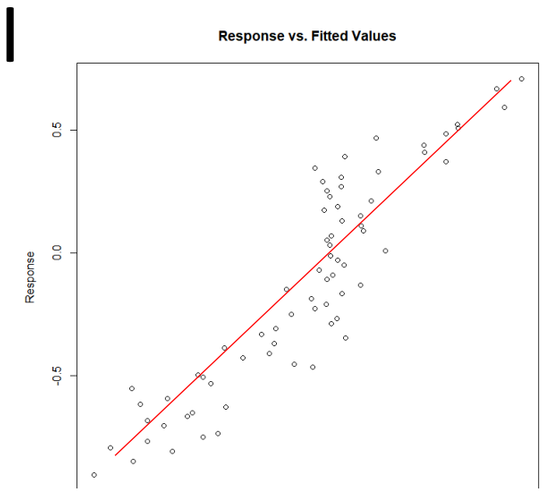
C

Reduced Model (refer to Fig. 1E, ID)	# Configurations per KO	# Rxns	# PCs Used for GAM	# PCs Selected by *GAMRandomStep()* Algorithm (Step Size)	R-sq (Adj.) of Fit, and Deviance Explained (of Modeled Data)
Minus Rounded same FVA Min & Max Reactions (3)	500	2912	526	?	
~~High Flux Backbone 20 (6)~~	~~500~~	~~981~~	~~526~~	~~?~~	
~~High Flux Backbone 50 (8)~~	~~500~~	~~981~~	~~526~~	~~?~~	
~~Minus Rounded same FVA Min & Max Reactions (3)~~	~~1 (Randomly chosen, Single FBA Solution)~~	~~2912~~	~~526~~	~~?~~	
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of 500 Configurations, unbinned)	2912	526	9 (5)	0.0588, 7.01%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of 500 Configurations, unbinned)	2912	526	7 (25)	0.126, 16.6%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of 500 Configurations, unbinned)	2912	526	8(25)	0.138, 18.8%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 528 bins)	2912	526	5 (20)	0.0503, 6.2%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 161 bins)	2912	526	9 (17)	0.463, 52.2%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 140 bins)	2912	526	11 (15)	0.503, 57.7%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 123 bins)	2912	526	8 (13)	0.59, 67.4%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points – 122 bins)	2912	526	12 (13)	0.566, 63.9%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 120 bins)	2912	526	11 (13)	0.569, 64.6%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 94 bins)	2912	526	7 (10)	0.679, 73.9%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 71 bins)	2912	526	5 (7)	0.743, 79.0%
Minus Rounded same FVA Min & Max Reactions (3)	1 (Centroid of Center Points - 71 bins)	2912	526	6 (7)	0.786, 83.8%

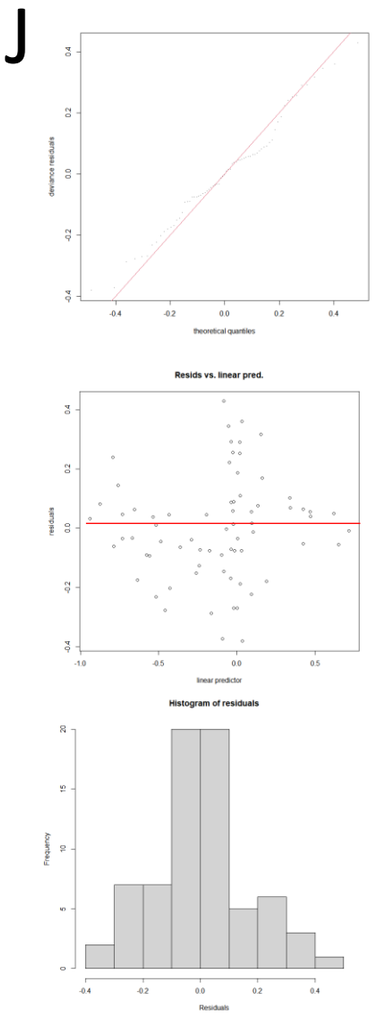
D



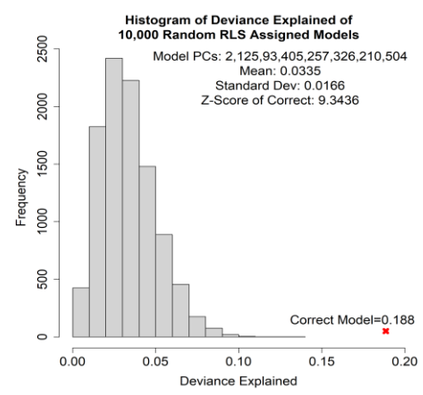
I



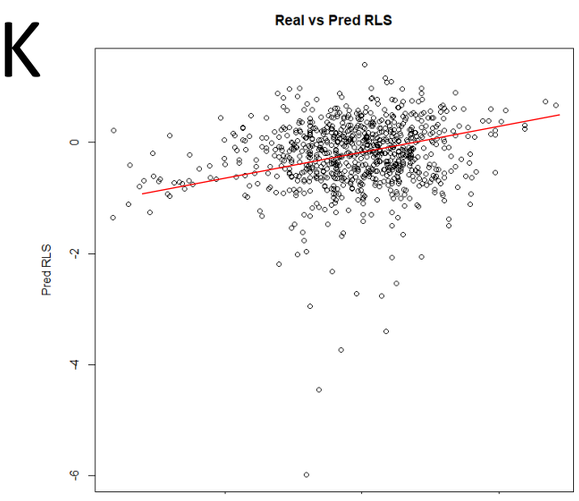
J



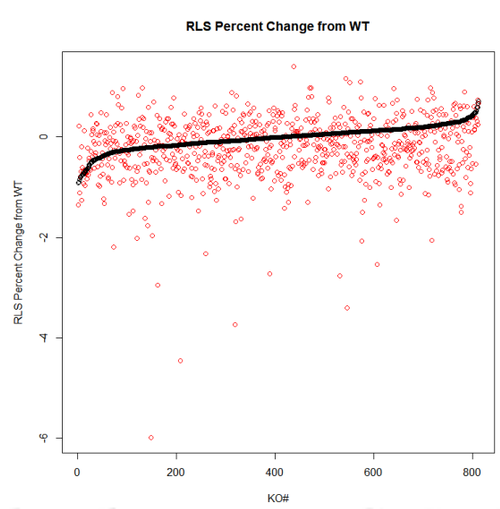
E



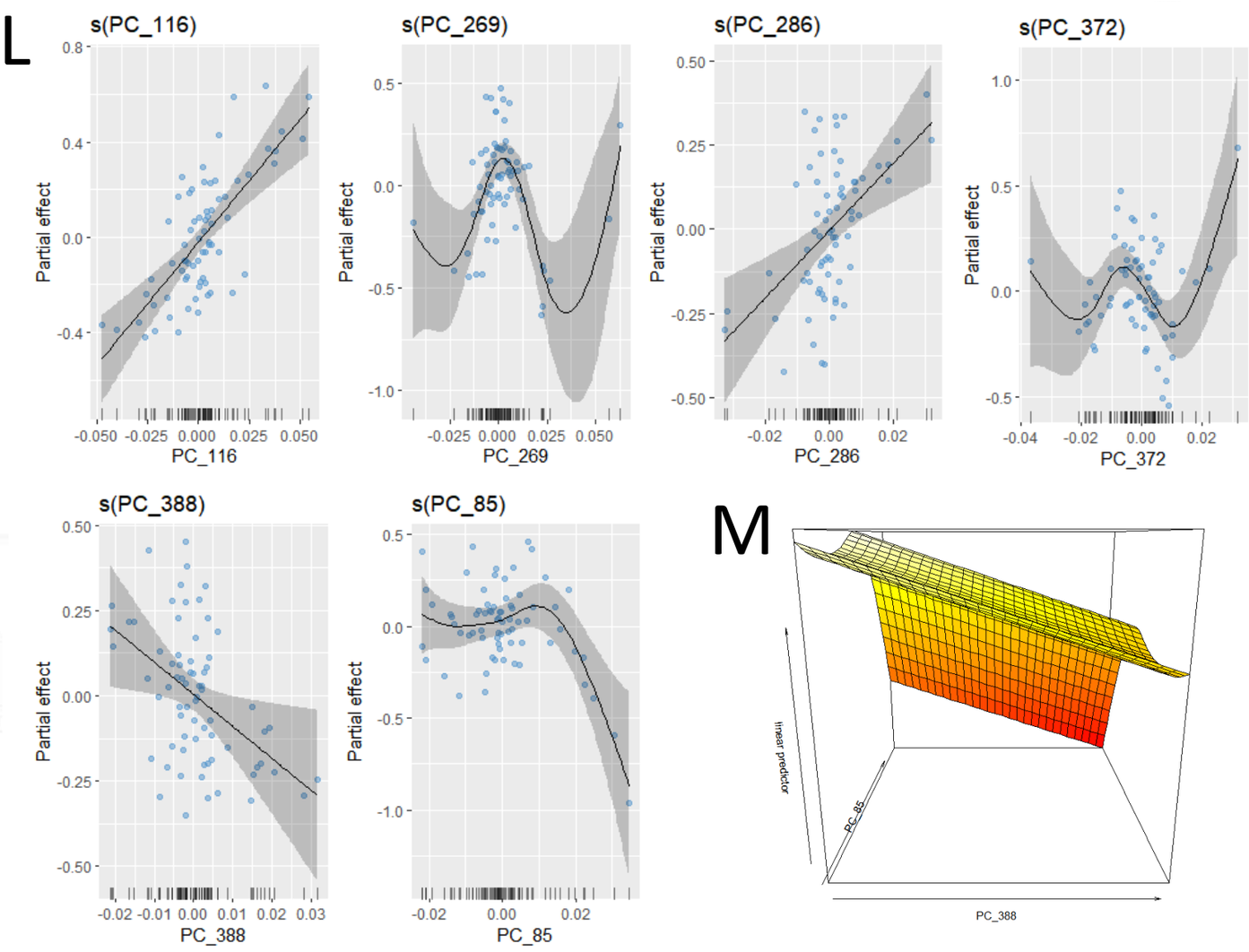
K



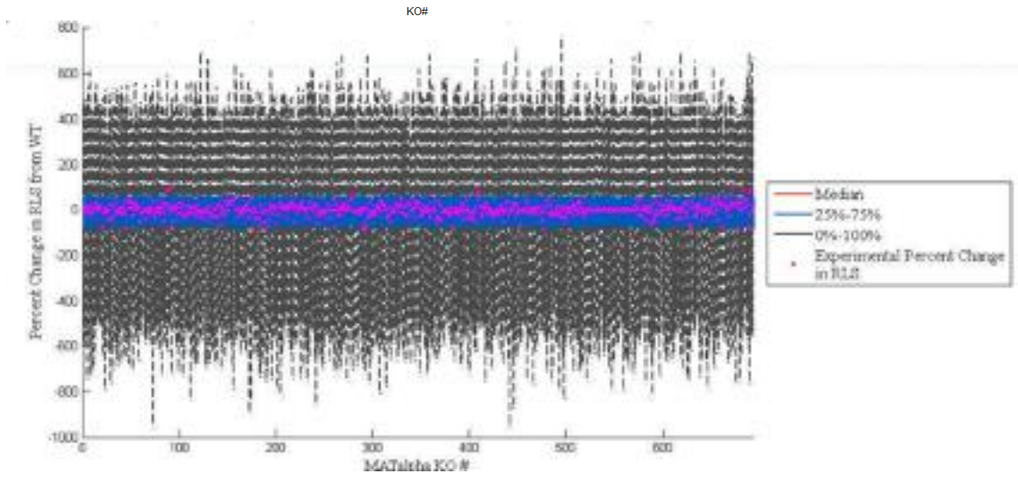
F



L

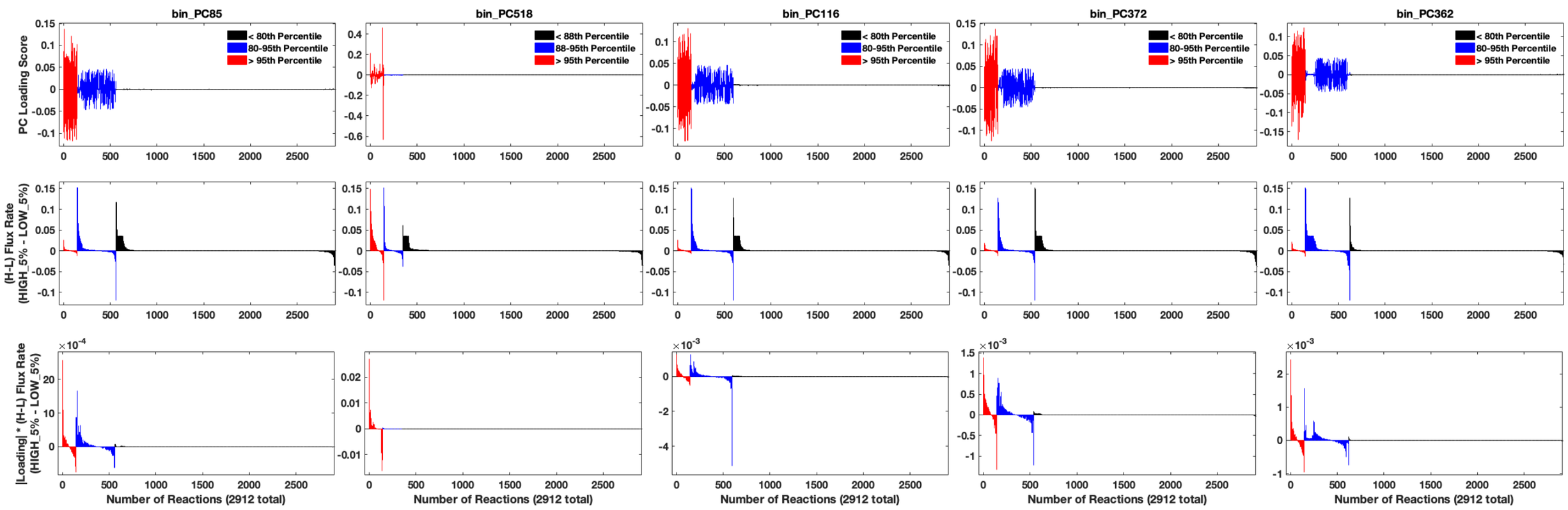


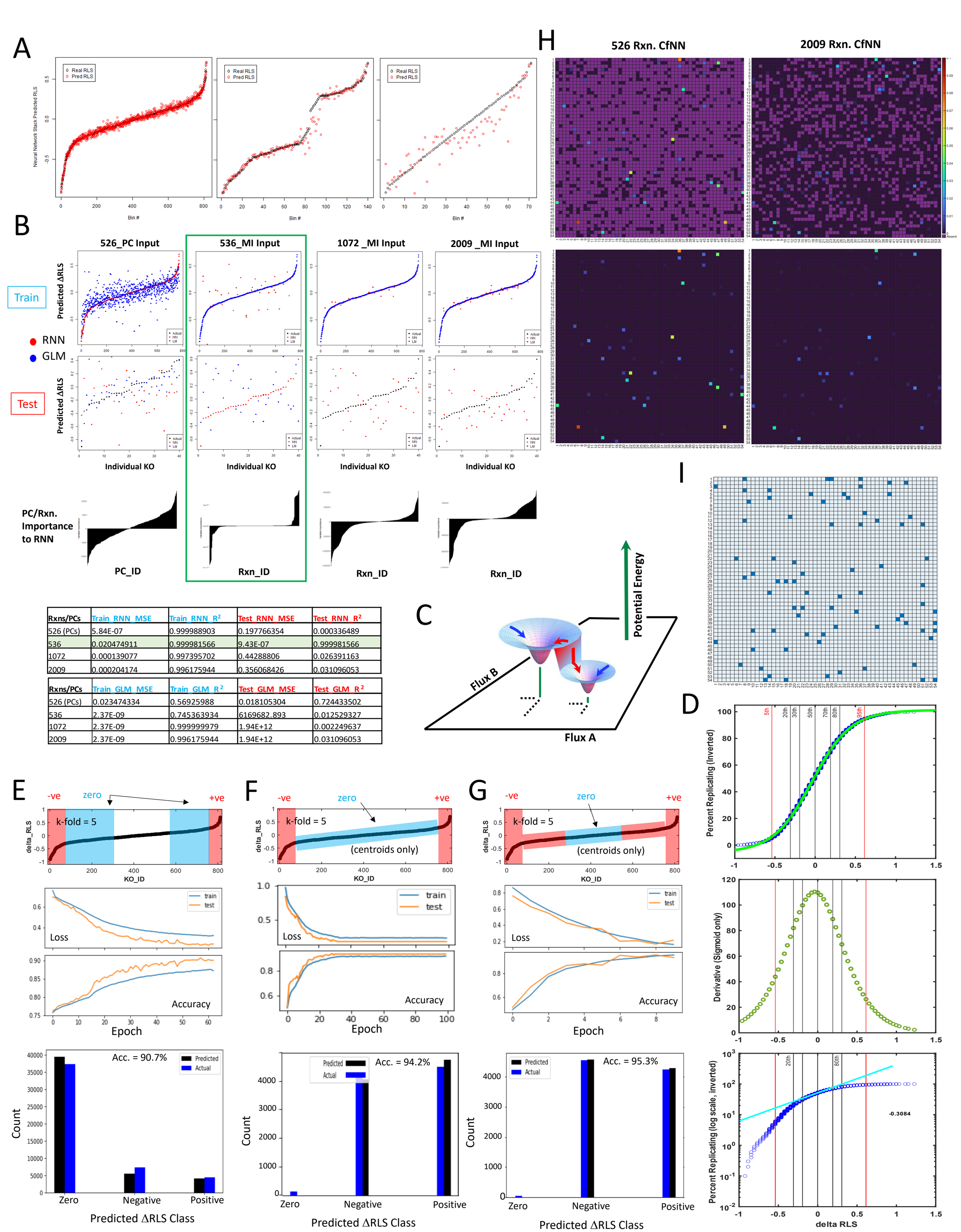
G



M

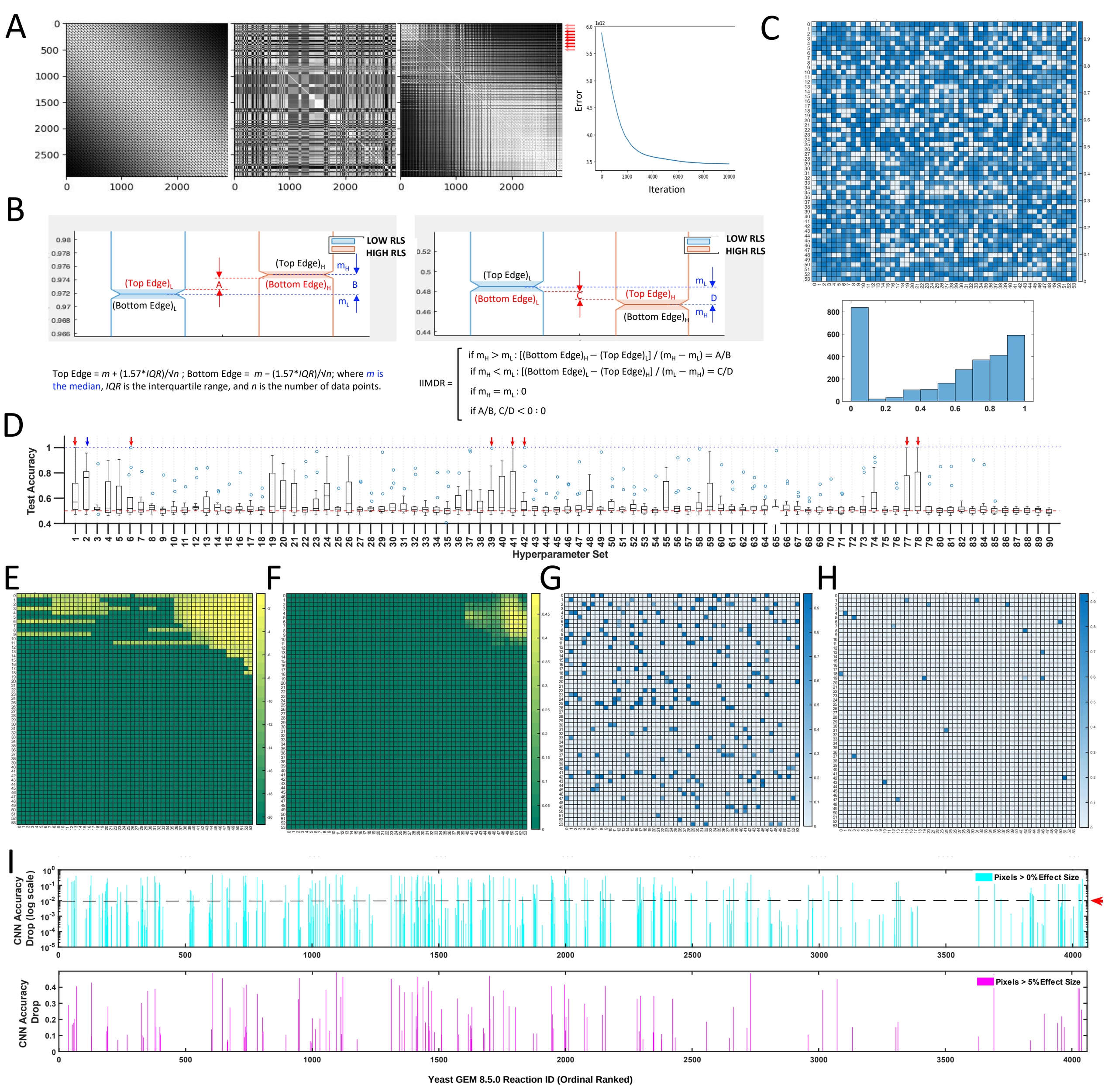
N

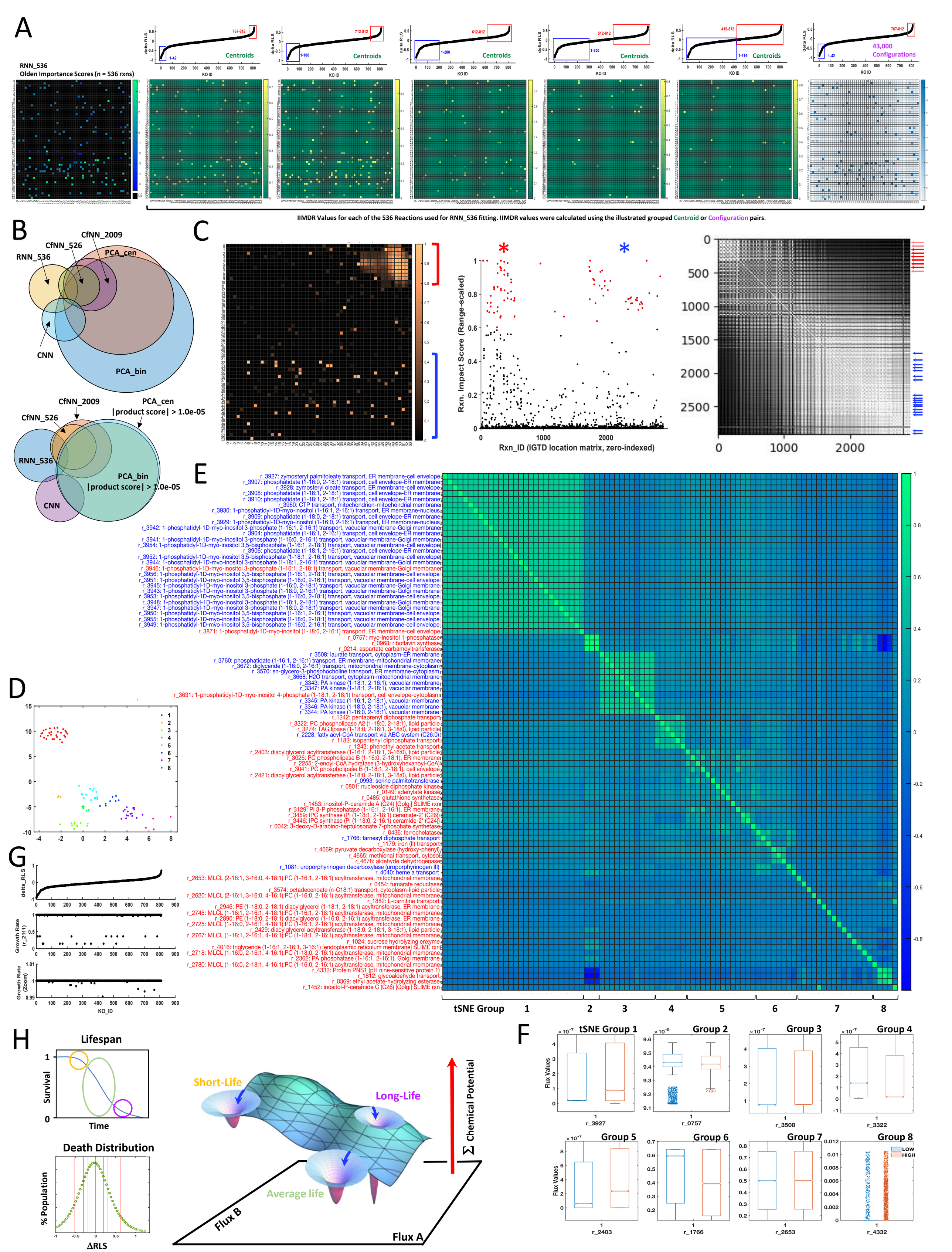




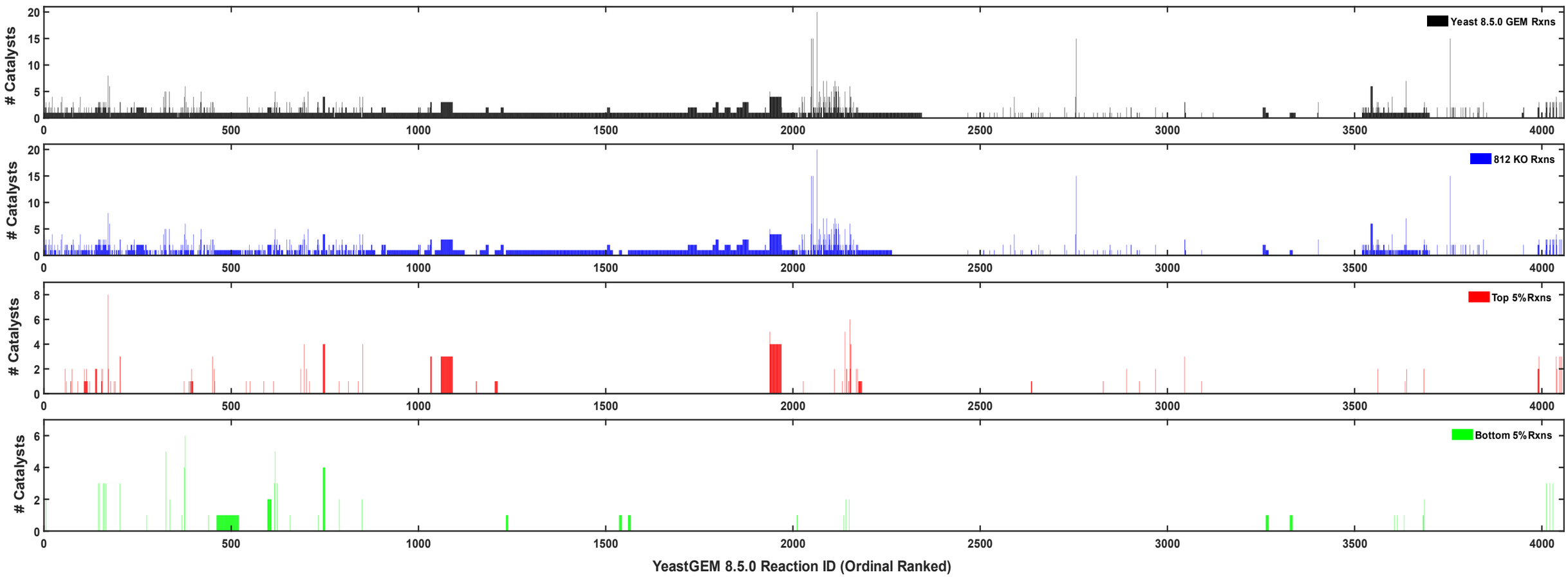
Rxns/PCs	Train_RNN_MSE	Train_RNN_R^2	Test_RNN_MSE	Test_RNN_R^2
526 (PCs)	5.84E-07	0.999988903	0.197766354	0.000336489
536	0.020474911	0.999981566	9.43E-07	0.999981566
1072	0.000139077	0.997395702	0.44288806	0.026391163
2009	0.000204174	0.996175944	0.356068426	0.031096053

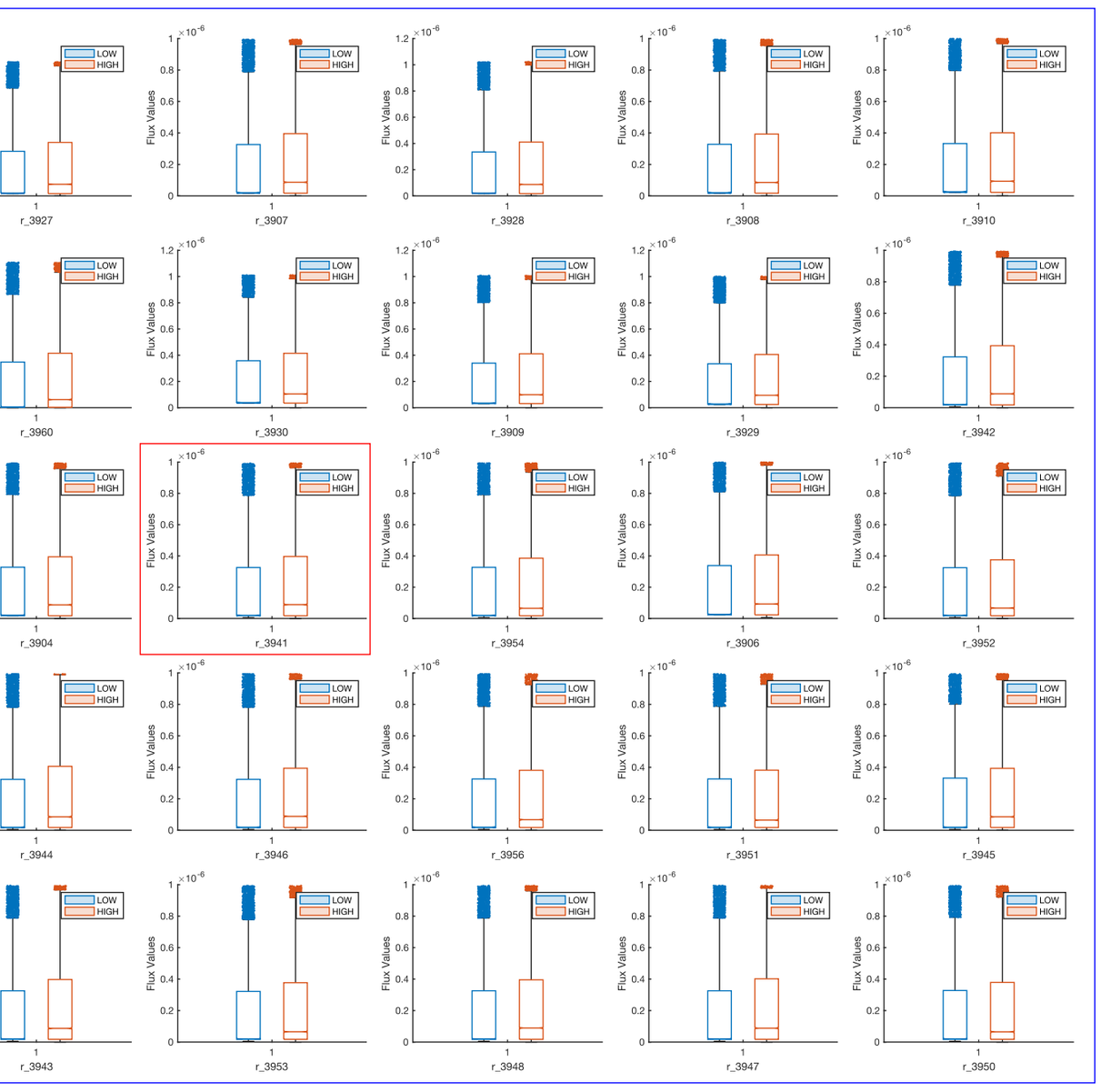
Rxns/PCs	Train_GLM_MSE	Train_GLM_R^2	Test_GLM_MSE	Test_GLM_R^2
526 (PCs)	0.023474334	0.56925988	0.018105304	0.724433502
536	2.37E-09	0.745363934	6169682.893	0.012529327
1072	2.37E-09	0.999999979	1.94E+12	0.002249637
2009	2.37E-09	0.996175944	1.94E+12	0.031096053

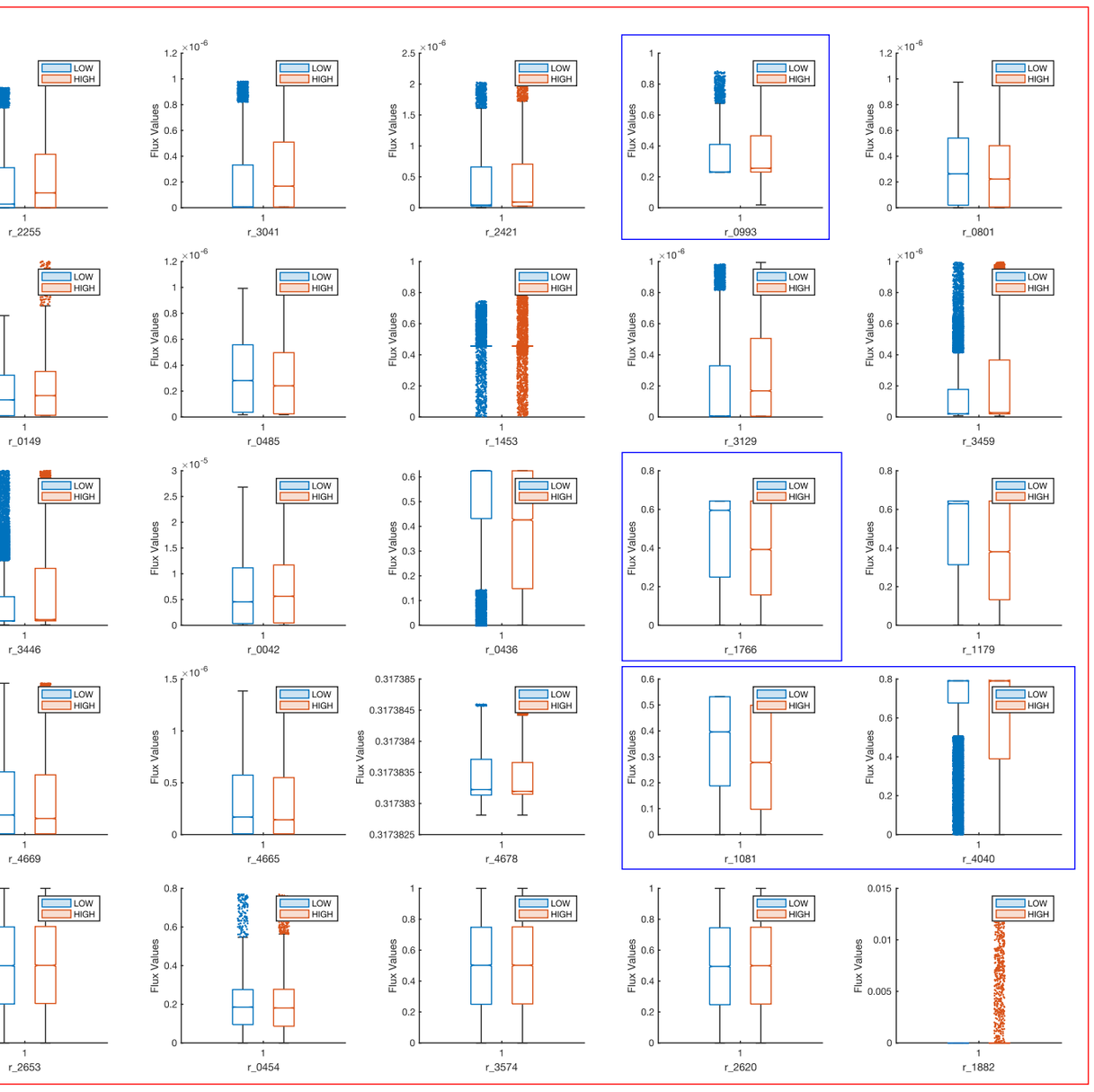

A
Error
Iteration
B
LOW RLS
HIGH RLS
(Top Edge)L
(Bottom Edge)L
(Top Edge)H
(Bottom Edge)H
mH
mL
A
B
C
D
Top Edge = m + (1.57*IQR)/√n ; Bottom Edge = m − (1.57*IQR)/√n; where m is the median, IQR is the interquartile range, and n is the number of data points.
IIMDR =
if mH > mL : [(Bottom Edge)H − (Top Edge)L] / (mH − mL) = A/B
if mH < mL : [(Bottom Edge)L − (Top Edge)H] / (mL − mH) = C/D
if mH = mL : 0
if A/B, C/D < 0 : 0
C
D
Test Accuracy
Hyperparameter Set
E
F
G
H
CNN Accuracy Drop (log scale)
Pixels > 0% Effect Size
CNN Accuracy Drop
Pixels > 5% Effect Size
Yeast GEM 8.5.0 Reaction ID (Ordinal Ranked)

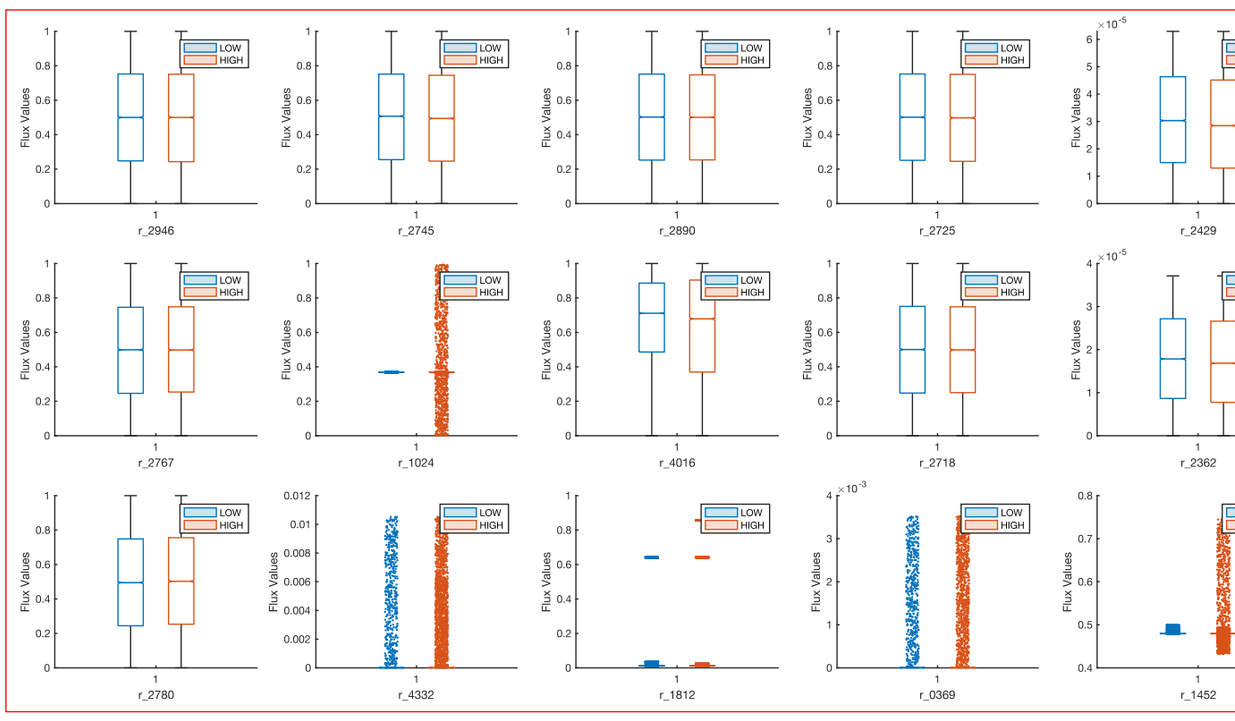

A
RNN_536
Olden Importance Scores (n = 536 rxns)
Centroids
43,000 Configurations
KO ID
delta RLS
IIMDR Values for each of the 536 Reactions used for RNN_536 fitting. IIMDR values were calculated using the illustrated grouped Centroid or Configuration pairs.
B
CfNN_2009
CfNN_526
RNN_536
PCA_cen
CNN
PCA_bin
|product score| > 1.0e-05
C
Rxn. Impact Score (Range-scaled)
Rxn_ID (IGTD location matrix, zero-indexed)
E
tSNE Group
D
G
delta_RLS
Growth Rate (r_2111)
Growth Rate (Zoom)
KO_ID
H
Lifespan
Survival
Time
Death Distribution
% Population
ΔRLS
Short-Life
Long-Life
Average life
Flux B
Flux A
Σ Chemical Potential
F
tSNE Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Flux Values
LOW
HIGH

A

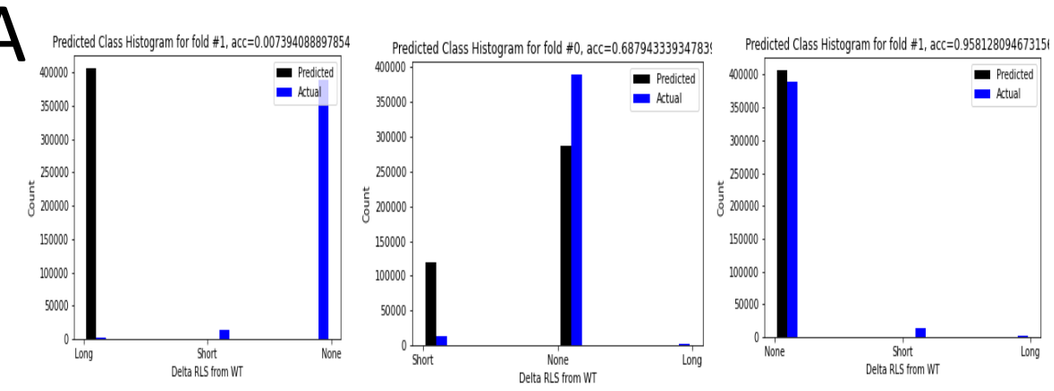

Yeast 8.5.0 GEM Rxns
812 KO Rxns
Top 5%Rxns
Bottom 5%Rxns
Catalysts
YeastGEM 8.5.0 Reaction ID (Ordinal Ranked)

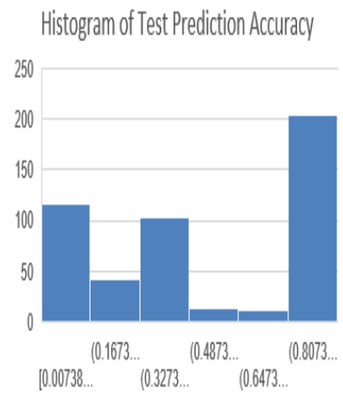

LOW
HIGH
Flux Values
r_3927
r_3907
r_3928
r_3908
r_3910
r_3960
r_3930
r_3909
r_3929
r_3942
r_3904
r_3941
r_3954
r_3906
r_3952
r_3944
r_3946
r_3956
r_3951
r_3945
r_3943
r_3953
r_3948
r_3947
r_3950

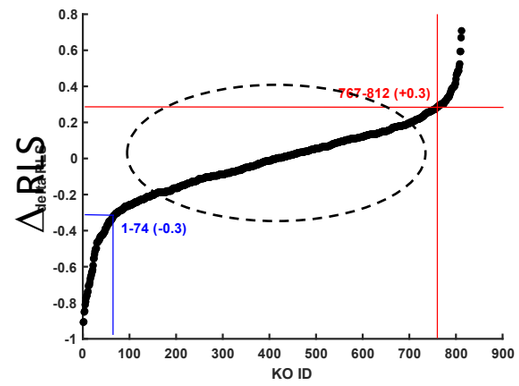

LOW
HIGH
Flux Values
r_2255
r_3041
r_2421
r_0993
r_0801
r_0149
r_0485
r_1453
r_3129
r_3459
r_3446
r_0042
r_0436
r_1766
r_1179
r_4669
r_4665
r_4678
r_1081
r_4040
r_2653
r_0454
r_3574
r_2620
r_1882


LOW
HIGH
Flux Values
r_2946
r_2745
r_2890
r_2725
r_2429
r_2767
r_1024
r_4016
r_2718
r_2362
r_2780
r_4332
r_1812
r_0369
r_1452

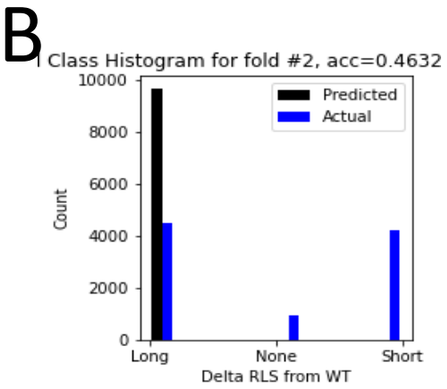
A

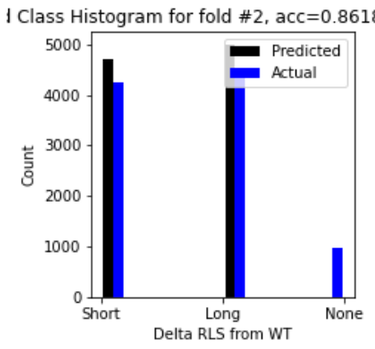


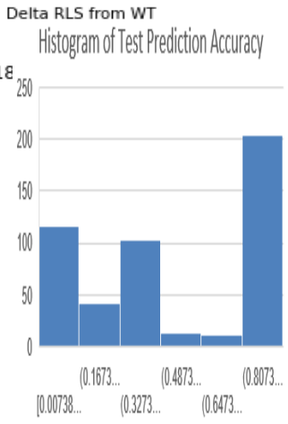




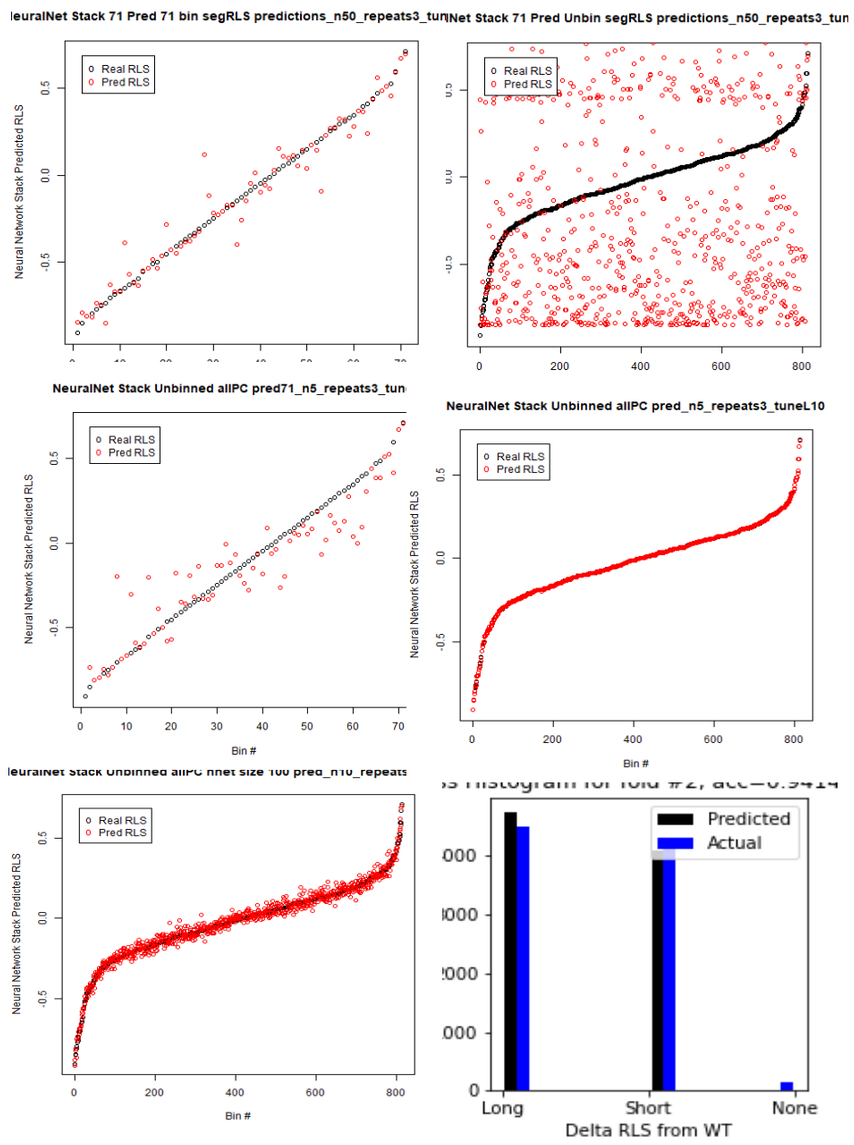
B







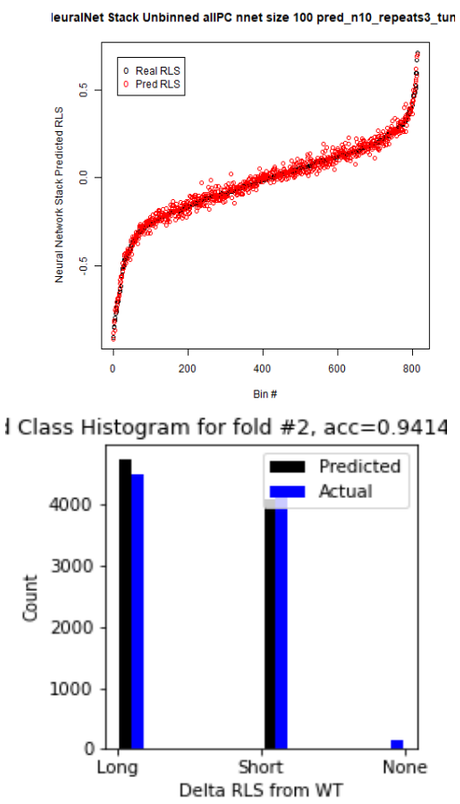
C



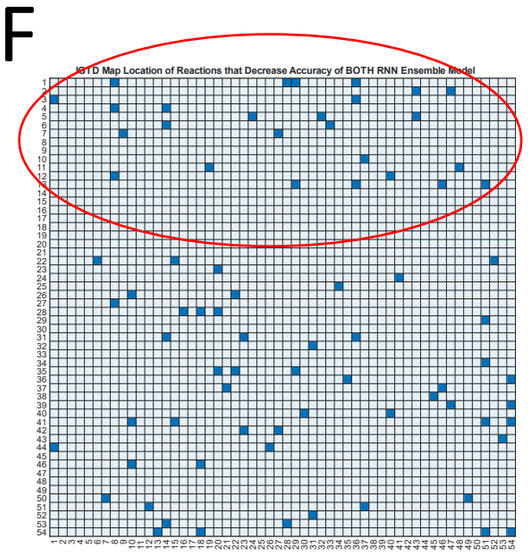
D

Batch Size	Drop Out Rate	Δ RLS Threshold	No. Input Neurons (Reactions)	No. folds	Training Accuracy (Average)	Training Accuracy (SDEV)	Test Accuracy (Average)	Test Accuracy (SDEV)
16	0	<|0.1|	537	5	0.899	0.015	0.897	0.015
16	0	<|0.3|	537	5	0.747	0.216	0.745	0.216
16	0.1	<|0.1|	537	5	0.832	0.006	0.831	0.006
16	0.1	<|0.3|	537	5	0.848	0.013	0.846	0.013
16	0.2	<|0.1|	537	5	0.747	0.030	0.747	0.030
16	0.2	<|0.3|	537	5	0.682	0.097	0.681	0.097
32	0	<|0.1|	537	5	0.862	0.045	0.859	0.045
32	0	<|0.3|	537	5	0.797	0.152	0.797	0.152
32	0.1	<|0.1|	537	5	0.812	0.056	0.812	0.056
32	0.1	<|0.3|	537	5	0.797	0.056	0.797	0.056
32	0.2	<|0.1|	537	5	0.662	0.123	0.660	0.123
32	0.2	<|0.3|	537	5	0.697	0.106	0.699	0.106
16	0	<|0.1|	2009	5	0.853	0.194	0.854	0.194
16	0	<|0.3|	2009	5	0.850	0.191	0.848	0.191
16	0.1	<|0.1|	2009	5	0.747	0.200	0.746	0.200
16	0.1	<|0.3|	2009	5	0.856	0.031	0.856	0.031
16	0.2	<|0.1|	2009	5	0.727	0.124	0.725	0.124
16	0.2	<|0.3|	2009	5	0.773	0.011	0.773	0.011
32	0	<|0.1|	2009	5	0.912	0.056	0.908	0.056
32	0	<|0.3|	2009	5	0.680	0.234	0.680	0.234
32	0.1	<|0.1|	2009	5	0.854	0.063	0.852	0.063
32	0.1	<|0.3|	2009	5	0.771	0.147	0.770	0.147
32	0.2	<|0.1|	2009	5	0.663	0.125	0.661	0.125
32	0.2	<|0.3|	2009	5	0.754	0.029	0.755	0.029
16	0	<|0.1|	2912	10	0.819	0.164	0.809	0.164
16	0	<|0.3|	2912	10	0.832	0.171	0.823	0.171
16	0.1	<|0.1|	2912	10	0.744	0.131	0.741	0.131
16	0.1	<|0.3|	2912	10	0.770	0.109	0.763	0.109
16	0.2	<|0.1|	2912	10	0.651	0.121	0.649	0.121
16	0.2	<|0.3|	2912	10	0.598	0.116	0.598	0.116
32	0	<|0.1|	2912	10	0.735	0.199	0.733	0.199
32	0	<|0.3|	2912	10	0.789	0.158	0.783	0.158
32	0.1	<|0.1|	2912	10	0.784	0.114	0.779	0.114
32	0.1	<|0.3|	2912	10	0.762	0.107	0.756	0.107
32	0.2	<|0.1|	2912	10	0.646	0.115	0.645	0.115
32	0.2	<|0.3|	2912	10	0.676	0.113	0.673	0.113

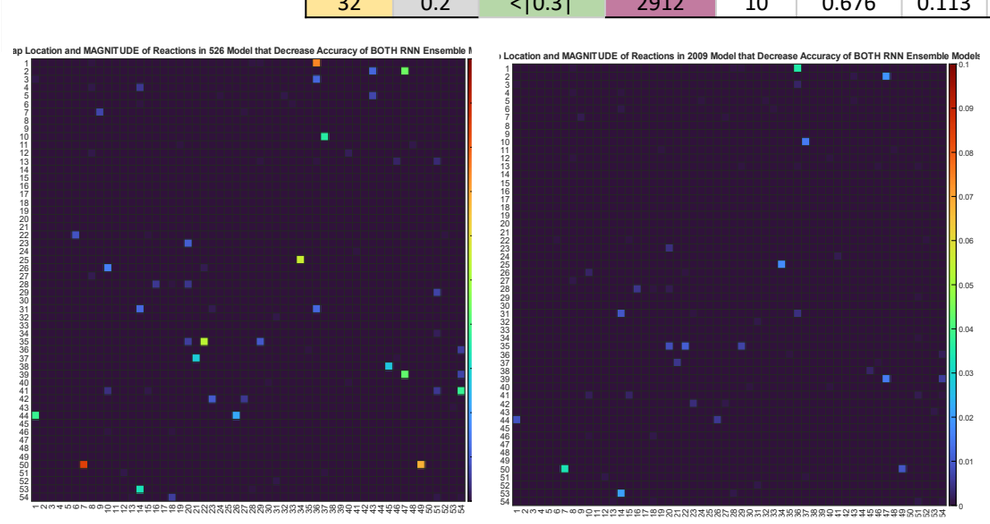
E



F



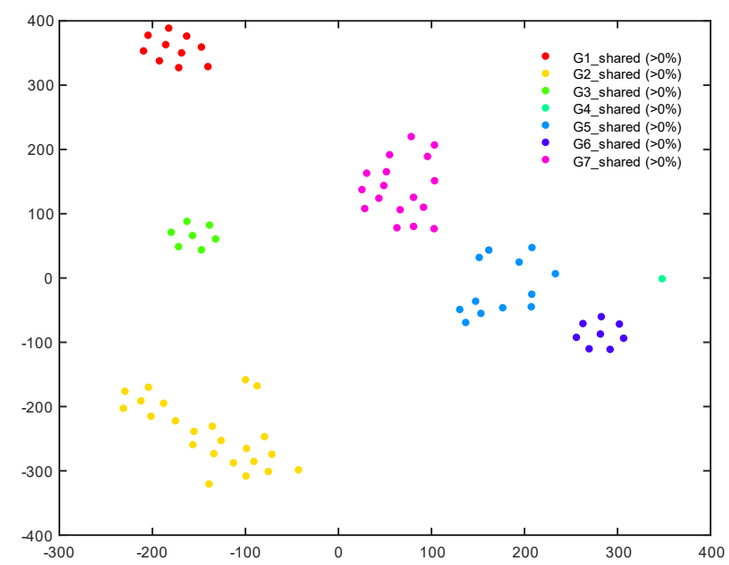
H



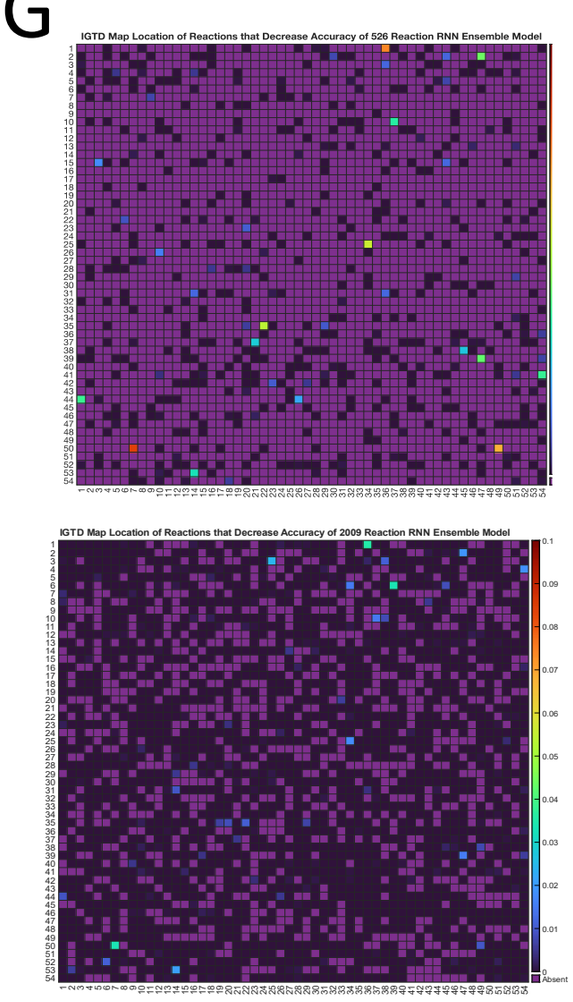
14.82%

3.88%

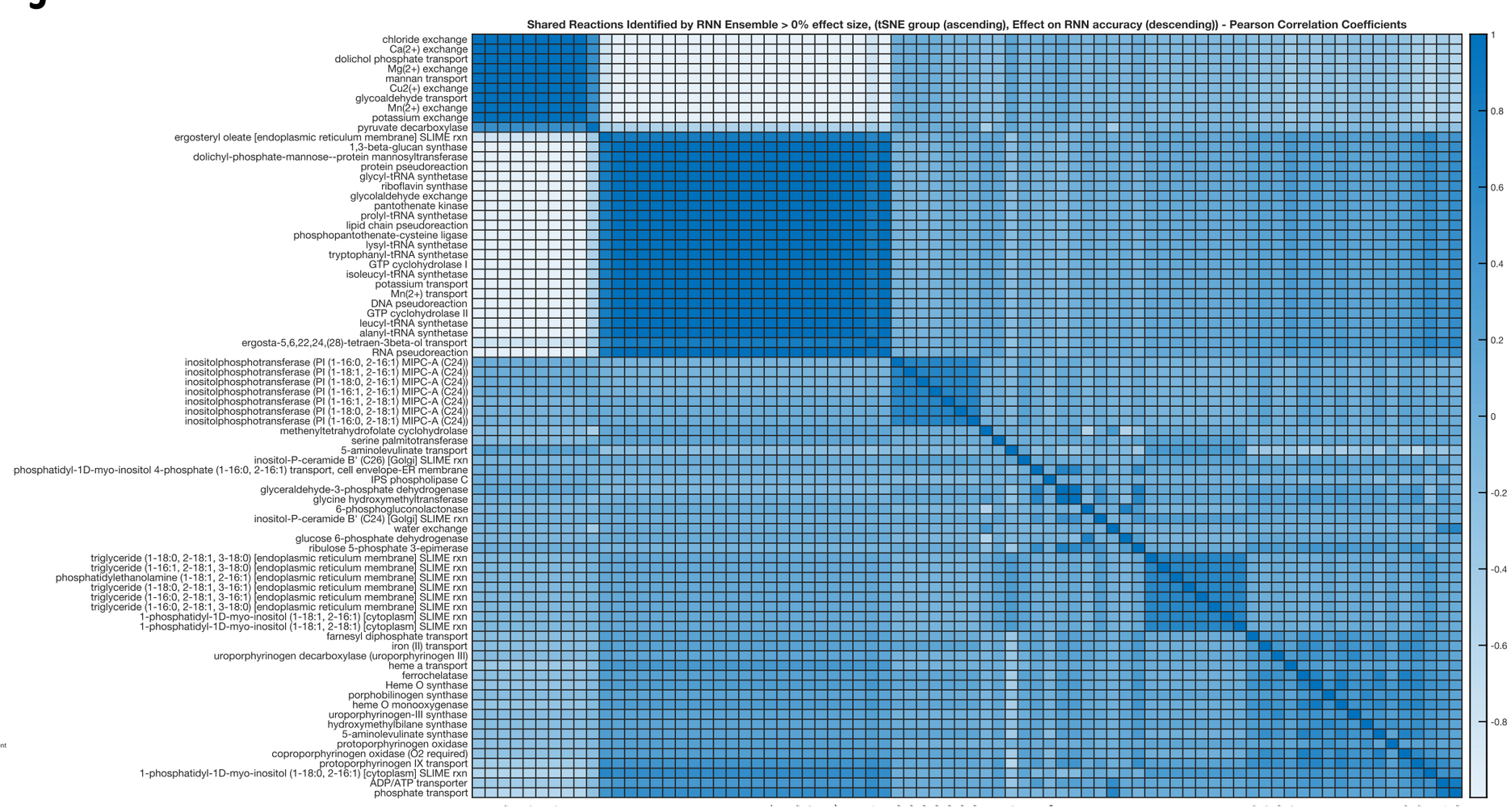
I

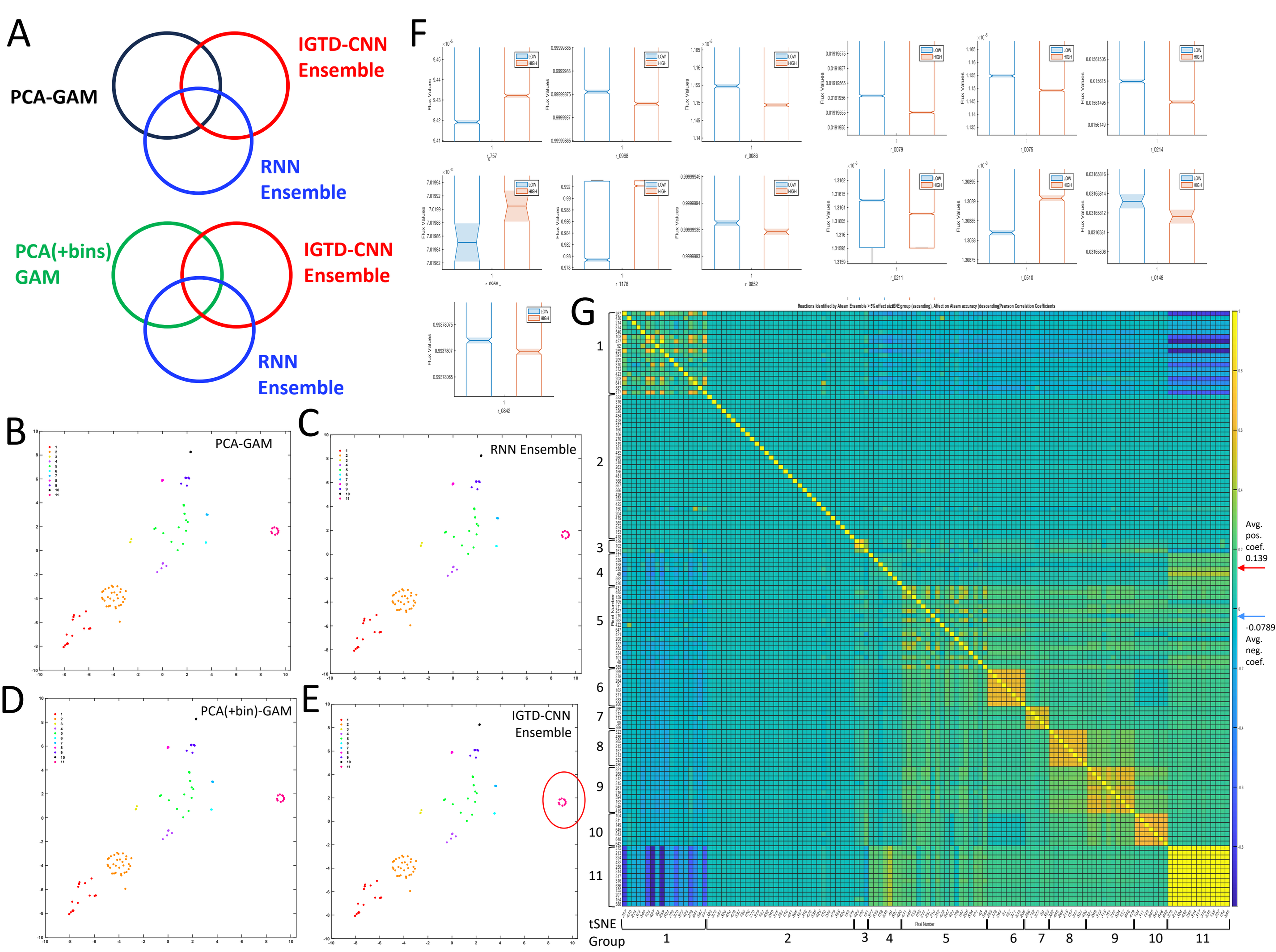


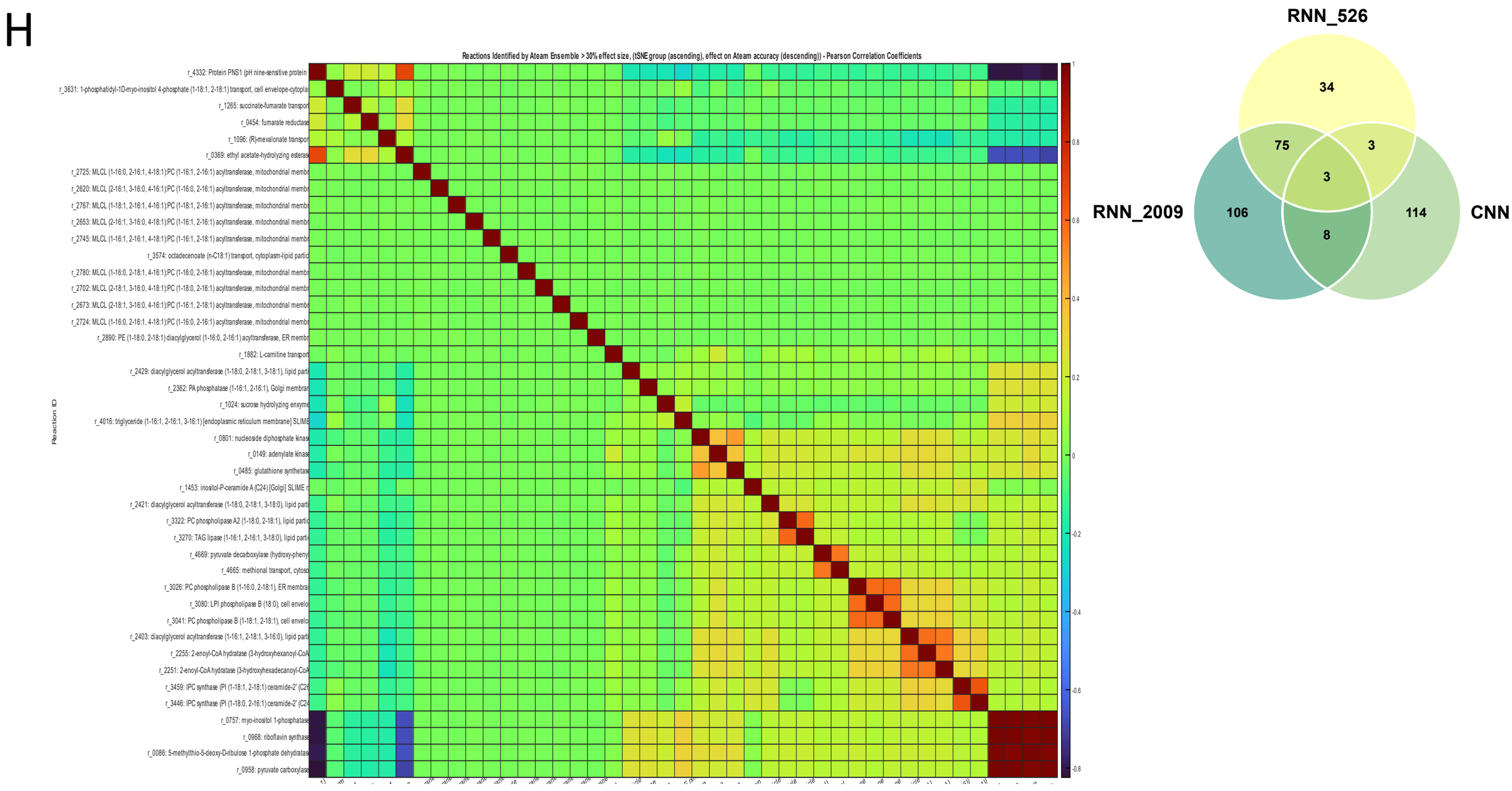
G

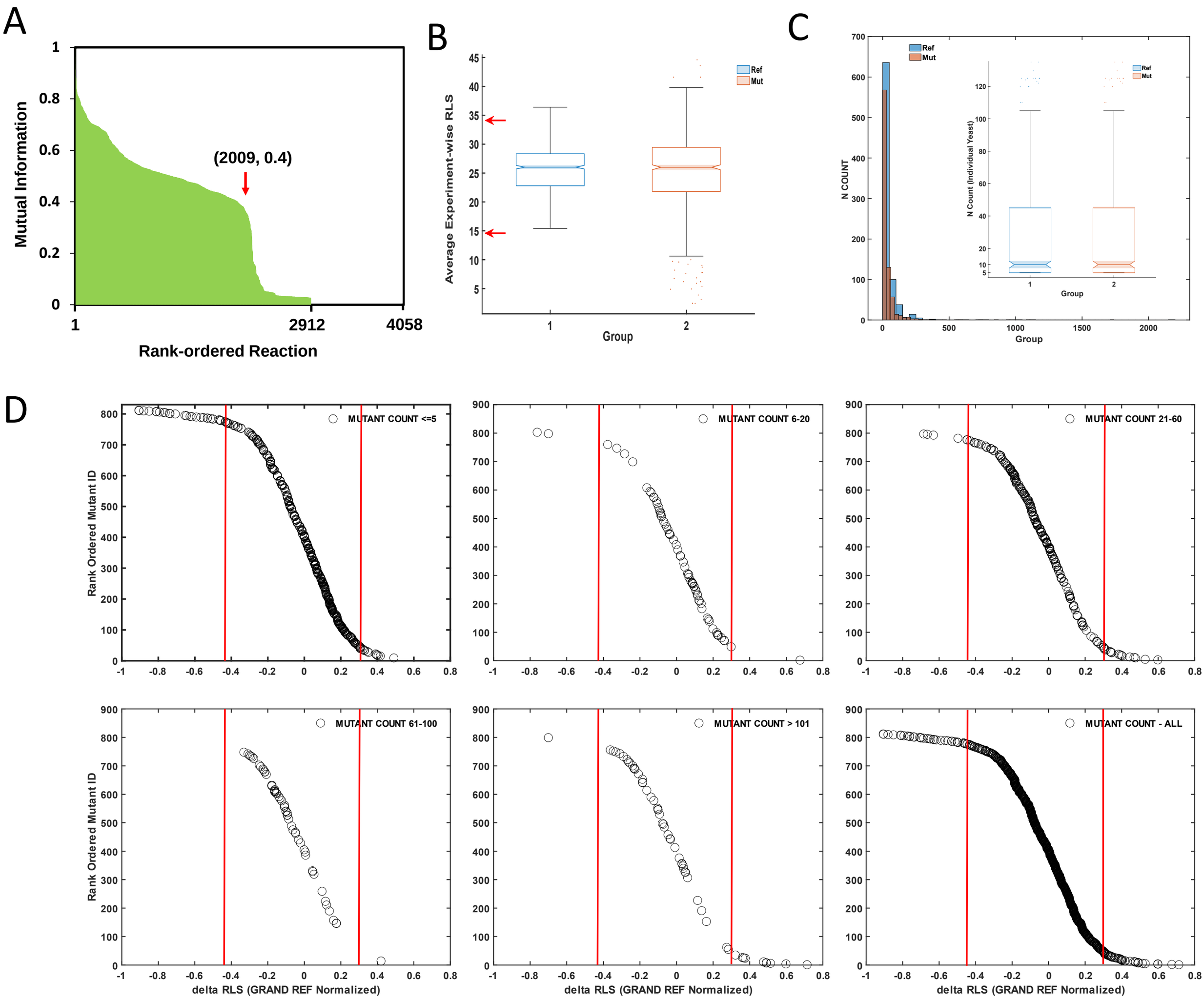


J

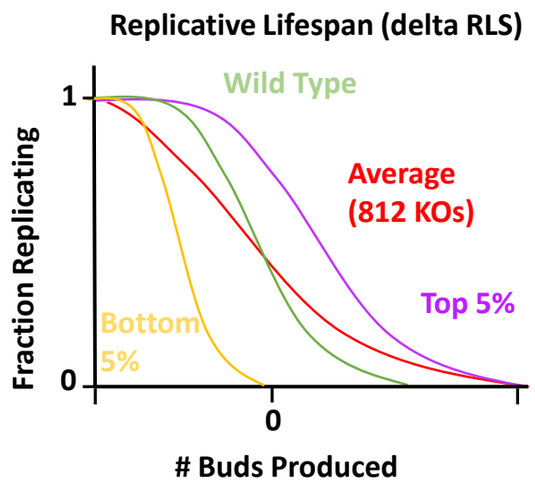



A
PCA-GAM
IGTD-CNN Ensemble
RNN Ensemble
PCA(+bins) GAM
IGTD-CNN Ensemble
RNN Ensemble
B
PCA-GAM
C
RNN Ensemble
D
PCA(+bin)-GAM
E
IGTD-CNN Ensemble
F
G
Avg. pos. coef. 0.139
-0.0789 Avg. neg. coef.
tSNE Group
1 2 3 4 5 6 7 8 9 10 11

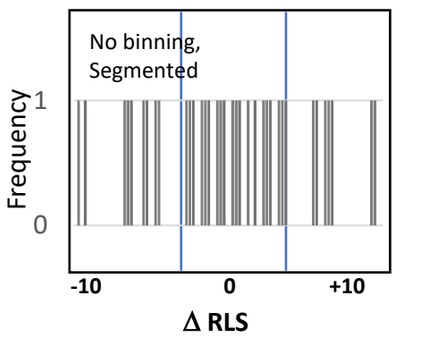

H
Reaction ID
RNN_526
34
75
3
3
RNN_2009
106
114
CNN
8

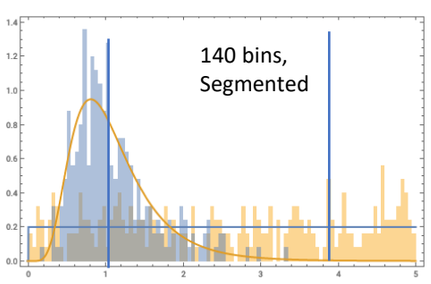

A
Mutual Information
(2009, 0.4)
Rank-ordered Reaction
B
Average Experiment-wise RLS
Ref
Mut
Group
C
N COUNT
N Count (Individual Yeast)
Group
D
Rank Ordered Mutant ID
MUTANT COUNT <=5
MUTANT COUNT 6-20
MUTANT COUNT 21-60
MUTANT COUNT 61-100
MUTANT COUNT > 101
MUTANT COUNT - ALL
delta RLS (GRAND REF Normalized)

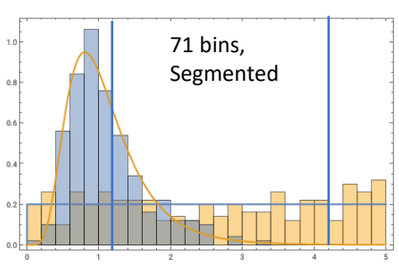
H

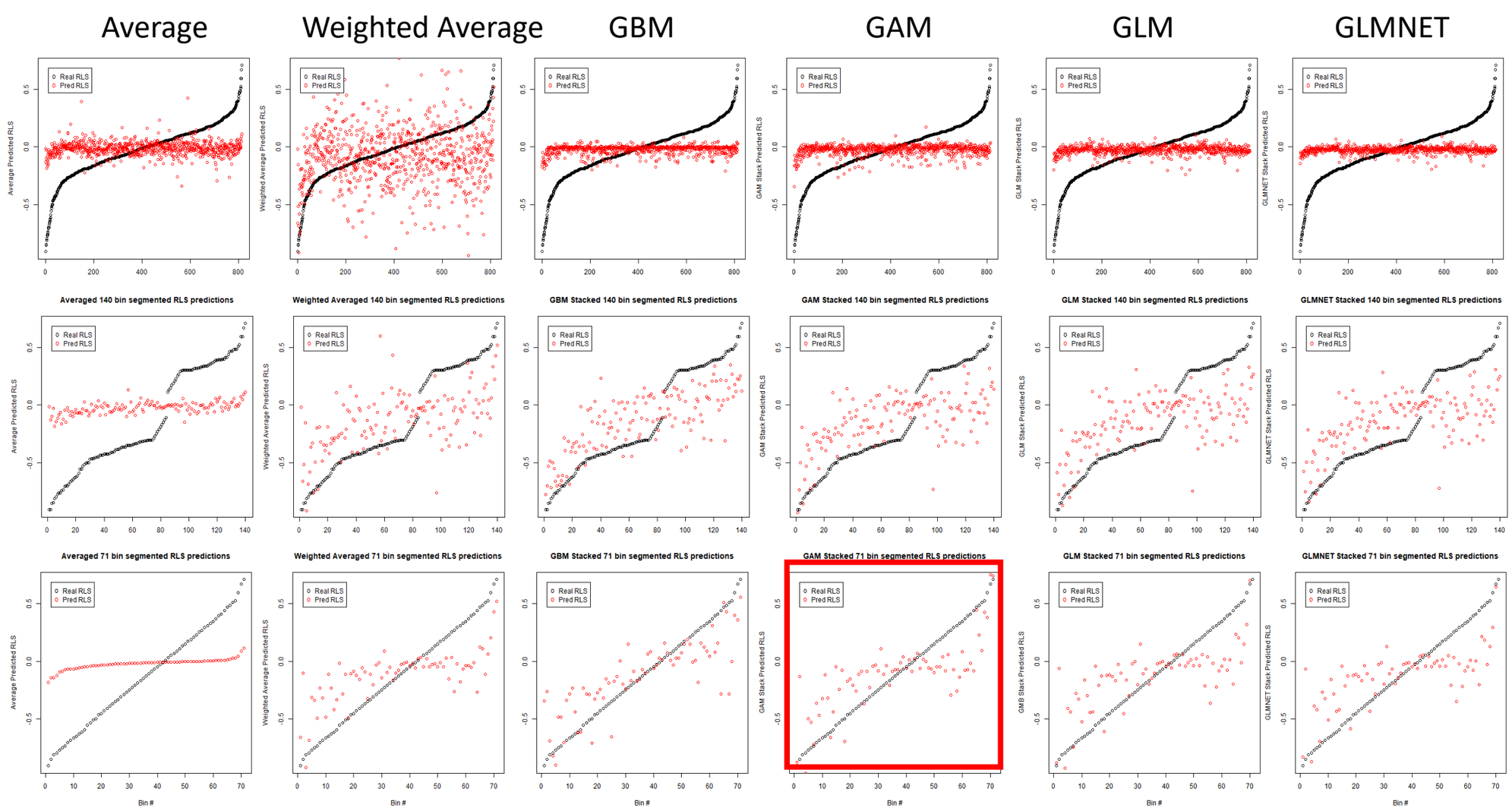

Replicative Lifespan (delta RLS)
Wild Type
Average
(812 KOs)
Top 5%
Bottom
5%
Fraction Replicating
1
0
0
Buds Produced

G

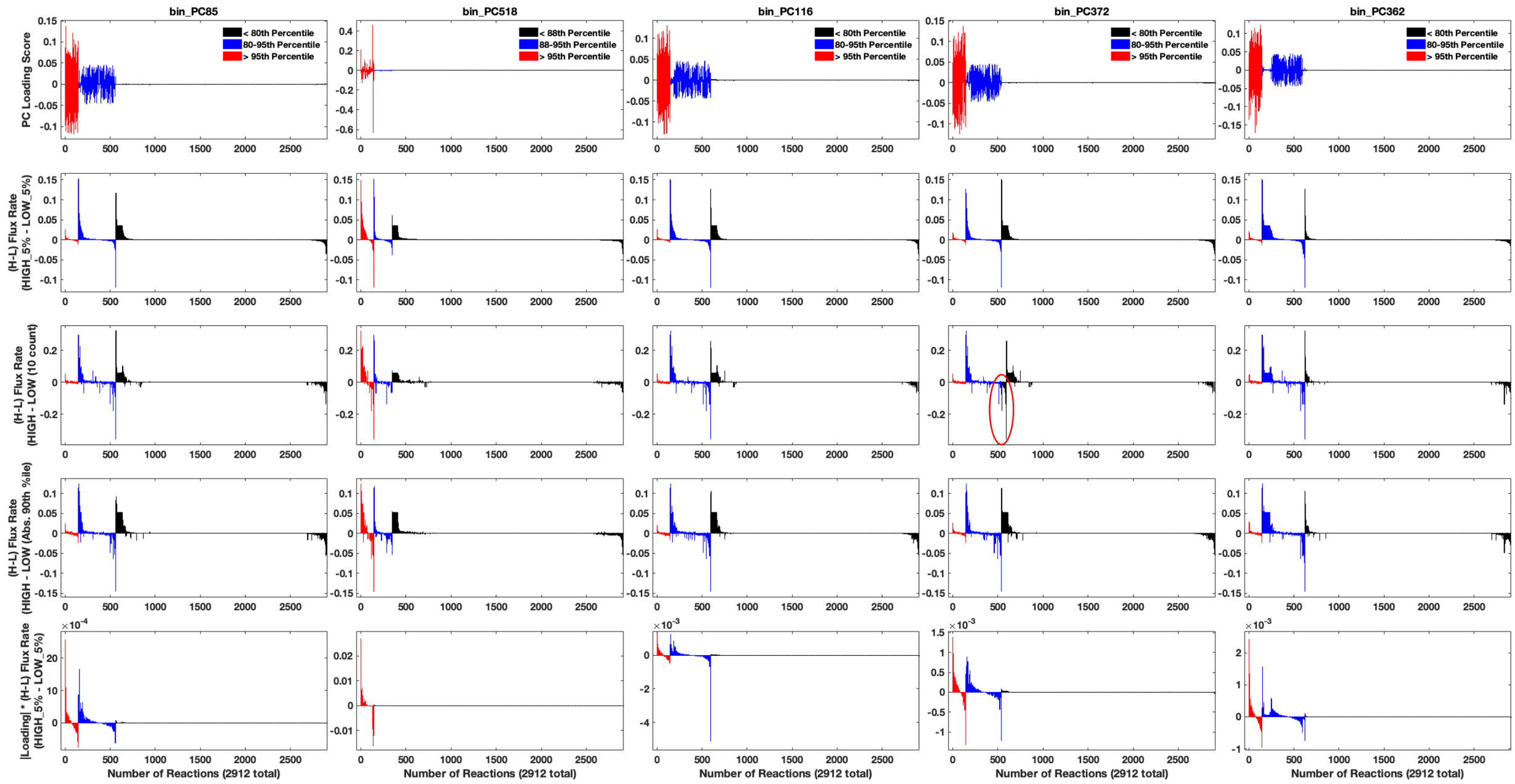

No binning,
Segmented
Frequency
1
0
-10
0
+10
Δ RLS


140 bins,
Segmented

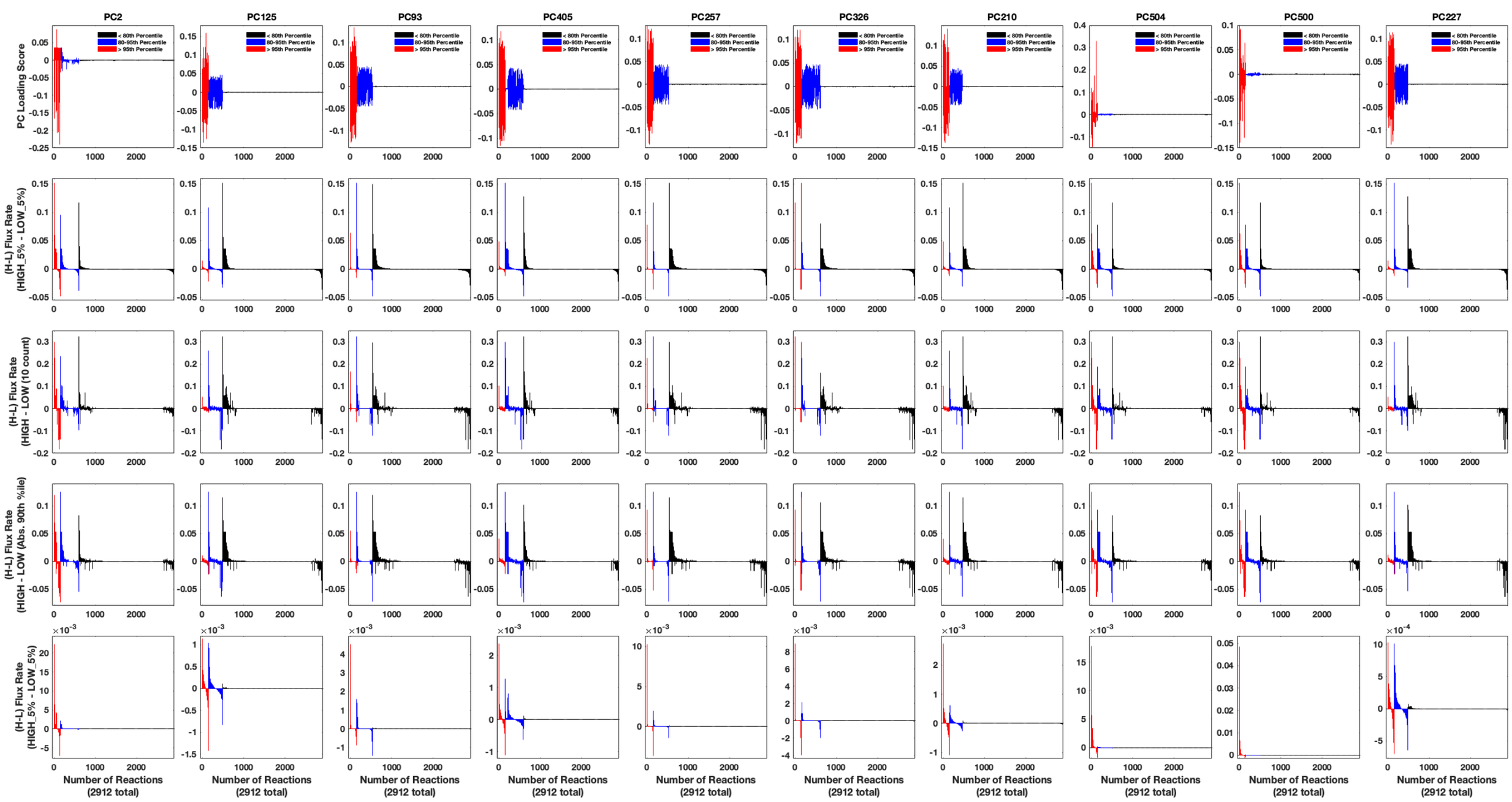

71 bins,
Segmented

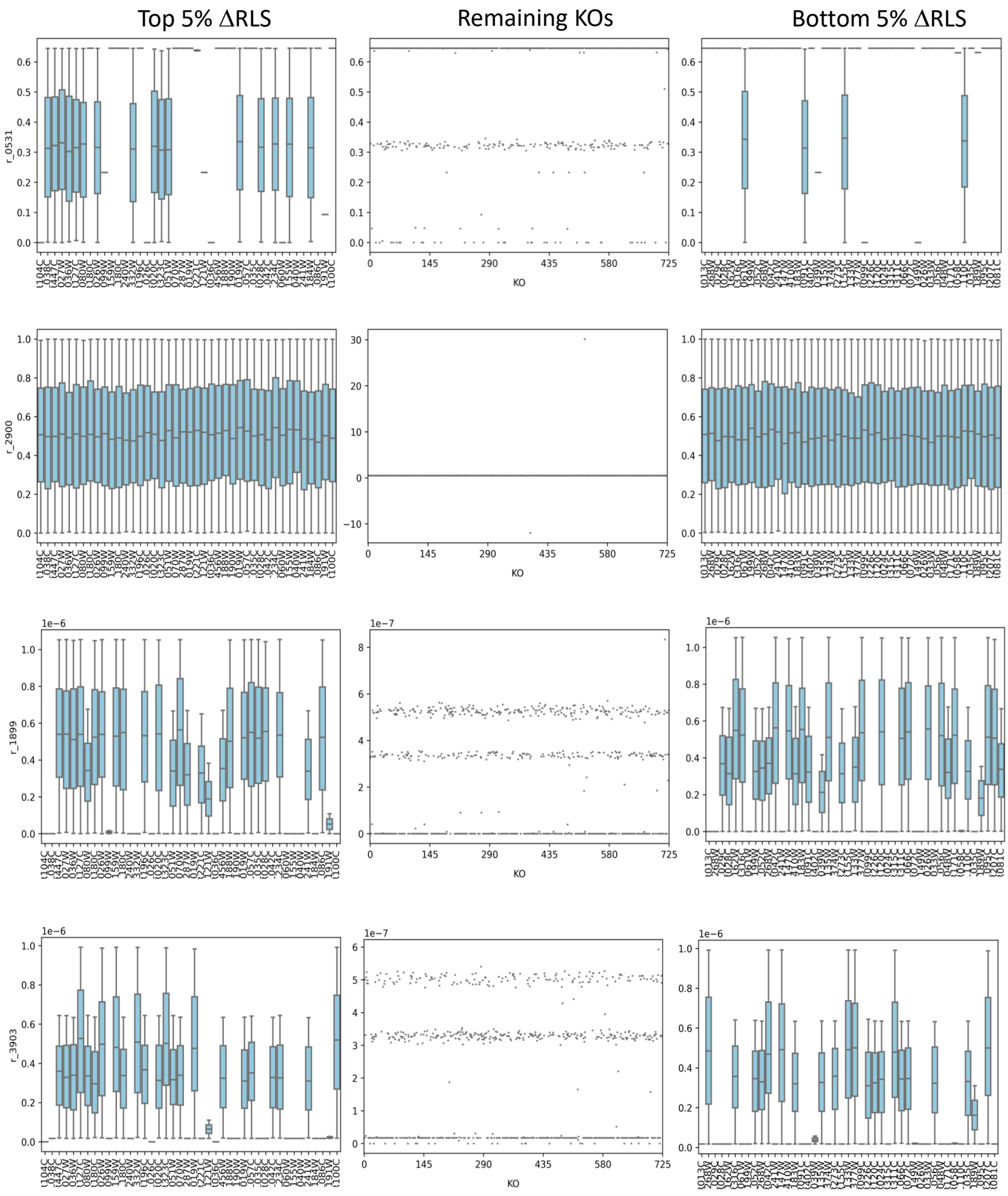

Average
Weighted Average
GBM
GAM
GLM
GLMNET
Real RLS
Pred RLS
Averaged 140 bin segmented RLS predictions
Weighted Averaged 140 bin segmented RLS predictions
GBM Stacked 140 bin segmented RLS predictions
GAM Stacked 140 bin segmented RLS predictions
GLM Stacked 140 bin segmented RLS predictions
GLMNET Stacked 140 bin segmented RLS predictions
Averaged 71 bin segmented RLS predictions
Weighted Averaged 71 bin segmented RLS predictions
GBM Stacked 71 bin segmented RLS predictions
GAM Stacked 71 bin segmented RLS predictions
GLM Stacked 71 bin segmented RLS predictions
GLMNET Stacked 71 bin segmented RLS predictions
Bin #

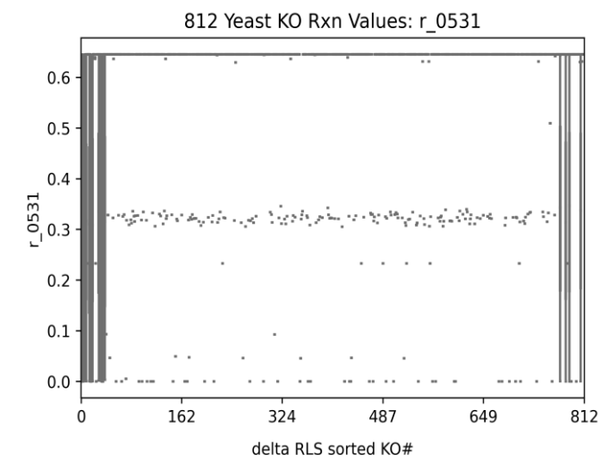
A

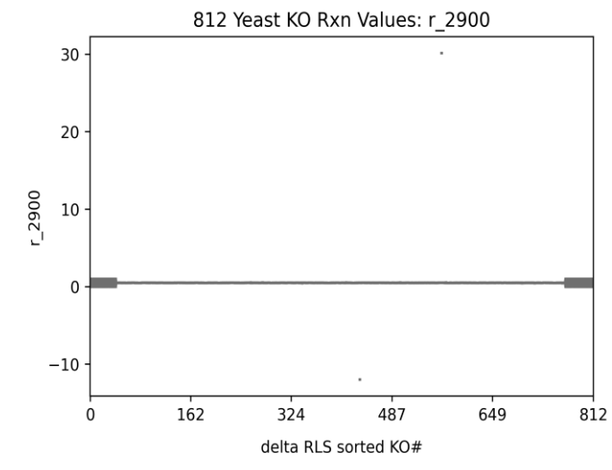

bin_PC85
bin_PC518
bin_PC116
bin_PC372
bin_PC362
< 80th Percentile
80-95th Percentile
> 95th Percentile
< 88th Percentile
88-95th Percentile
PC Loading Score
(H-L) Flux Rate (HIGH_5% - LOW_5%)
(H-L) Flux Rate (HIGH - LOW (10 count)
(H-L) Flux Rate (HIGH - LOW (Abs. 90th %ile)
|Loading| * (H-L) Flux Rate (HIGH_5% - LOW_5%)
Number of Reactions (2912 total)

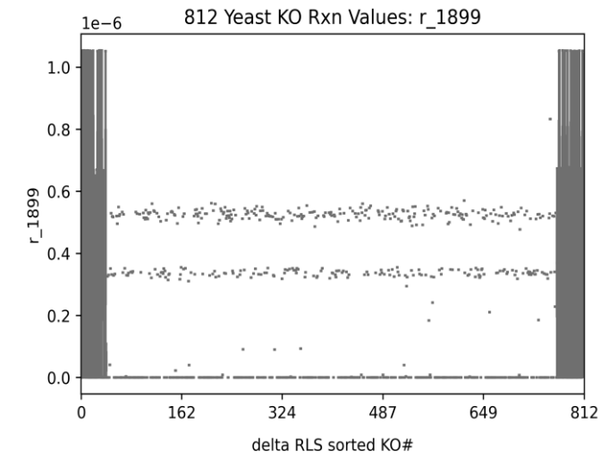
B

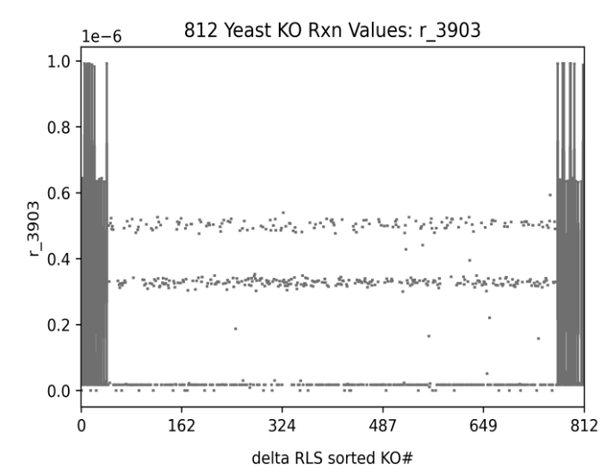

PC2
PC125
PC93
PC405
PC257
PC326
PC210
PC504
PC500
PC227
< 80th Percentile
80-95th Percentile
> 95th Percentile
PC Loading Score
(H-L) Flux Rate (HIGH_5% - LOW_5%)
(H-L) Flux Rate (HIGH - LOW (10 count)
(H-L) Flux Rate (HIGH - LOW (Abs. 90th %ile)
(H-L) Flux Rate (HIGH_5% - LOW_5%)
Number of Reactions (2912 total)

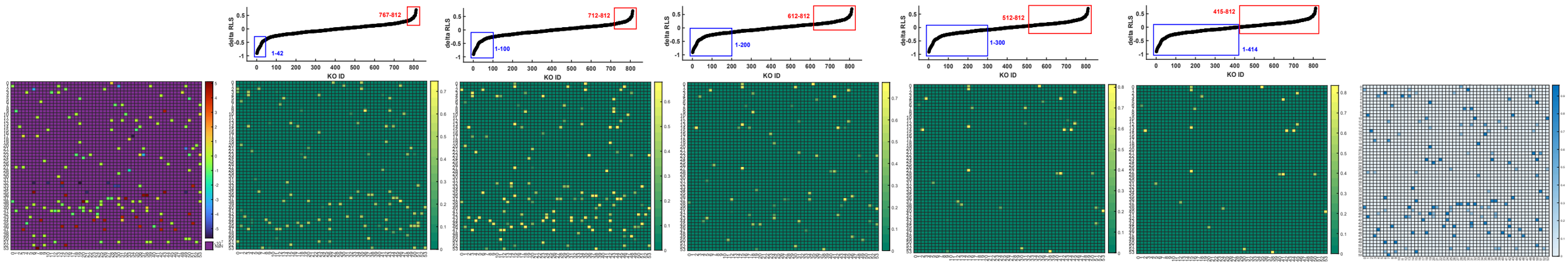

Top 5% ΔRLS
Remaining KOs
Bottom 5% ΔRLS
r_0531
r_2900
r_1899
r_3903
KO
1e−6
1e−7

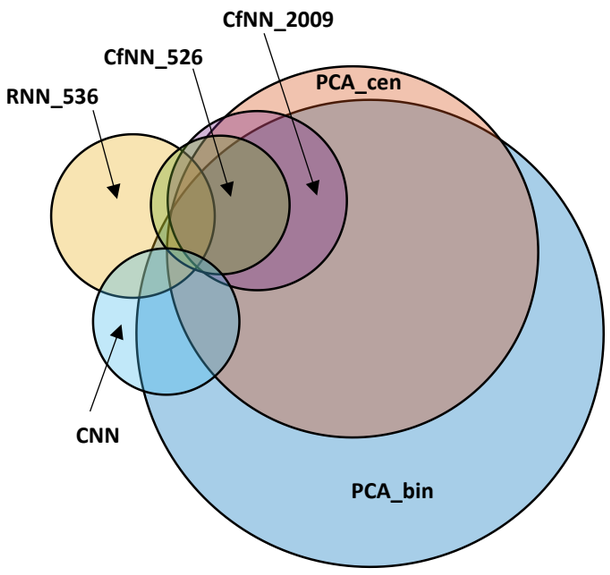

812 Yeast KO Rxn Values: r_0531
r_0531
delta RLS sorted KO#

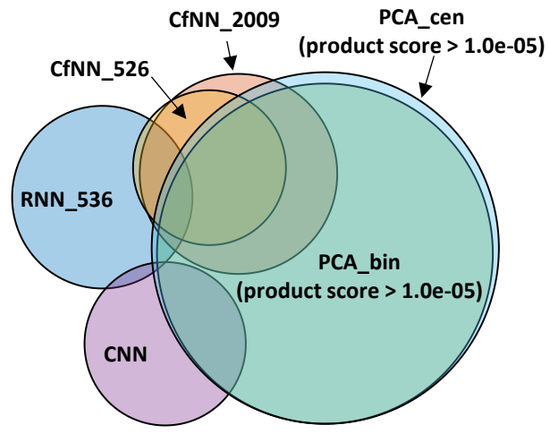

812 Yeast KO Rxn Values: r_2900
r_2900
delta RLS sorted KO#

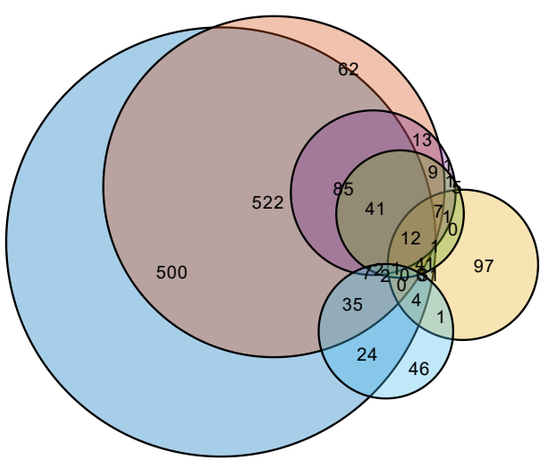

812 Yeast KO Rxn Values: r_1899
1e−6
r_1899
delta RLS sorted KO#

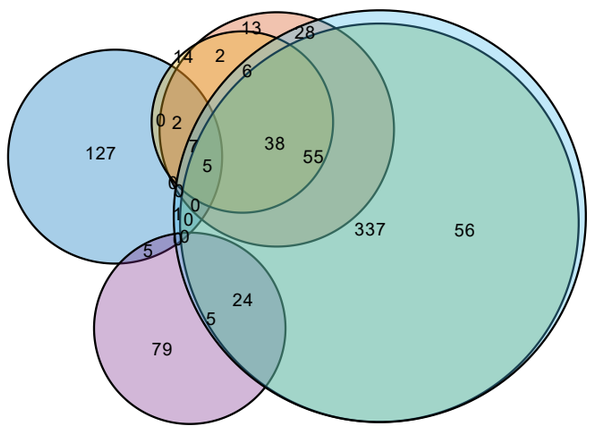

812 Yeast KO Rxn Values: r_3903
1e−6
r_3903
delta RLS sorted KO#

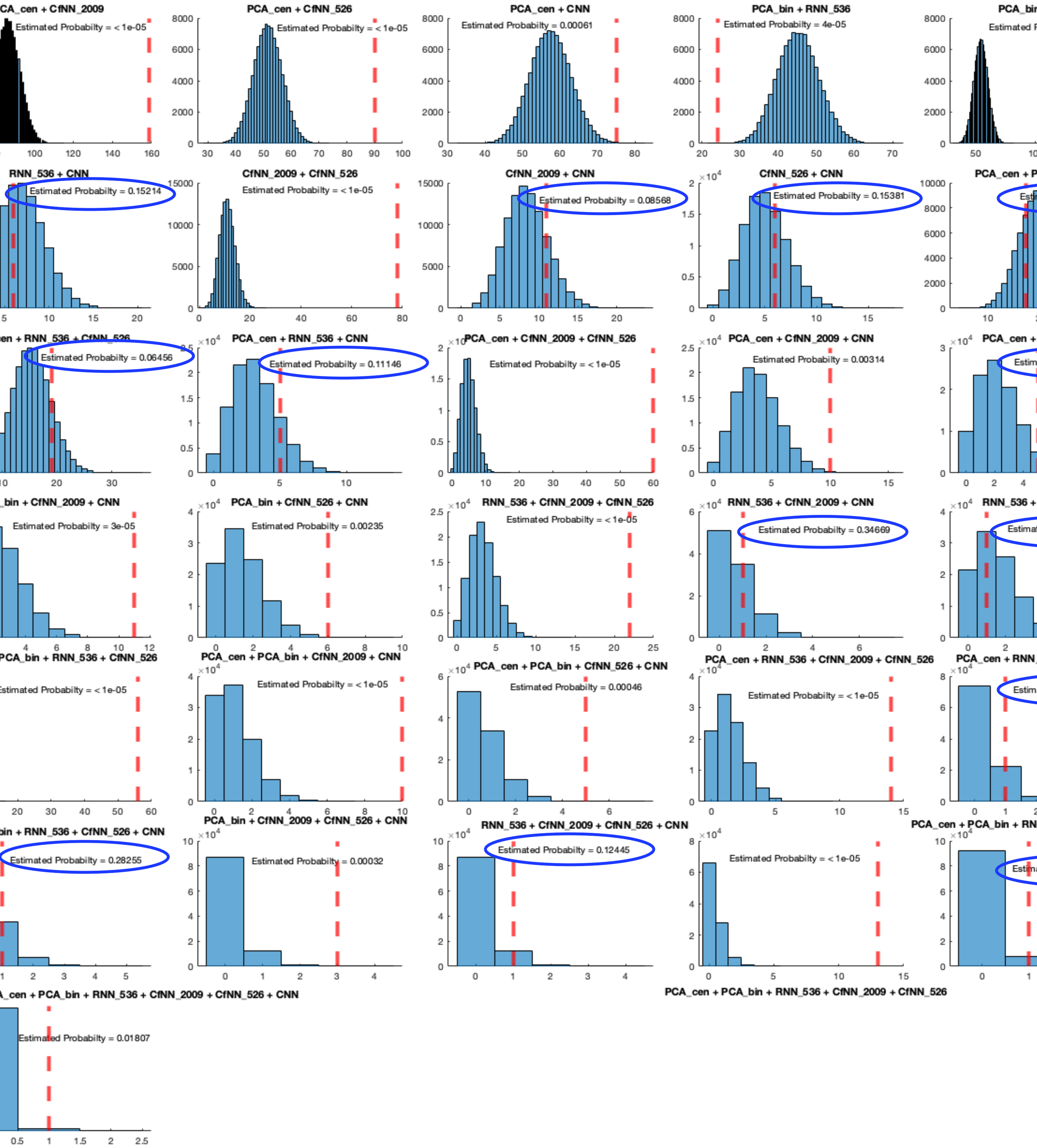

delta RLS
KO ID
767-812
1-42
712-812
1-100
612-812
1-200
512-812
1-300
415-812
1-414


CfNN_2009
CfNN_526
RNN_536
PCA_cen
CNN
PCA_bin


CfNN_2009
PCA_cen
(product score > 1.0e-05)
CfNN_526
RNN_536
PCA_bin
(product score > 1.0e-05)
CNN


62
13
9
522
85
41
12
500
97
35
4
1
24
46


13
28
14
2
6
2
38
55
127
5
337
56
24
5
79


CA_cen + CfNN_2009
Estimated Probabilty = < 1e-05
PCA_cen + CfNN_526
Estimated Probabilty = < 1e-05
PCA_cen + CNN
Estimated Probabilty = 0.00061
PCA_bin + RNN_536
Estimated Probabilty = 4e-05
PCA_bin
Estimated
RNN_536 + CNN
Estimated Probabilty = 0.15214
CfNN_2009 + CfNN_526
Estimated Probabilty = < 1e-05
CfNN_2009 + CNN
Estimated Probabilty = 0.08568
CfNN_526 + CNN
Estimated Probabilty = 0.15381
PCA_cen + P
Estimated
en + RNN_536 + CfNN_526
Estimated Probabilty = 0.06456
PCA_cen + RNN_536 + CNN
Estimated Probabilty = 0.11146
PCA_cen + CfNN_2009 + CfNN_526
Estimated Probabilty = < 1e-05
PCA_cen + CfNN_2009 + CNN
Estimated Probabilty = 0.00314
PCA_cen +
Estim
_bin + CfNN_2009 + CNN
Estimated Probabilty = 3e-05
PCA_bin + CfNN_526 + CNN
Estimated Probabilty = 0.00235
RNN_536 + CfNN_2009 + CfNN_526
Estimated Probabilty = < 1e-05
RNN_536 + CfNN_2009 + CNN
Estimated Probabilty = 0.34669
RNN_536 +
Estimat
PCA_bin + RNN_536 + CfNN_526
Estimated Probabilty = < 1e-05
PCA_cen + PCA_bin + CfNN_2009 + CNN
Estimated Probabilty = < 1e-05
PCA_cen + PCA_bin + CfNN_526 + CNN
Estimated Probabilty = 0.00046
PCA_cen + RNN_536 + CfNN_2009 + CfNN_526
Estimated Probabilty = < 1e-05
PCA_cen + RNN
Estim
bin + RNN_536 + CfNN_526 + CNN
Estimated Probabilty = 0.28255
PCA_bin + CfNN_2009 + CfNN_526 + CNN
Estimated Probabilty = 0.00032
RNN_536 + CfNN_2009 + CfNN_526 + CNN
Estimated Probabilty = 0.12445
Estimated Probabilty = < 1e-05
PCA_cen + PCA_bin + RN
Estim
A_cen + PCA_bin + RNN_536 + CfNN_2009 + CfNN_526 + CNN
Estimated Probabilty = 0.01807
PCA_cen + PCA_bin + RNN_536 + CfNN_2009 + CfNN_526