Chronos: a CRISPR cell population dynamics model

Joshua M. Dempster[1], Isabella Boyle[1], Francisca Vazquez[1], David Root[1], Jesse S. Boehm[1], William C. Hahn[1,2], Aviad Tsherniak[1], James M. McFarland[1]

[1] Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA
[2] Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215, USA

Abstract

CRISPR loss of function screens are a powerful tool to interrogate cancer biology but are known to exhibit a number of biases and artifacts that can confound the results, such as DNA cutting toxicity, incomplete phenotype penetrance and screen quality bias. Computational methods that more faithfully model the CRISPR biological experiment could more effectively extract the biology of interest than typical current methods. Here we introduce Chronos, an algorithm for inferring gene knockout fitness effects based on an explicit model of the dynamics of cell proliferation after CRISPR gene knockout. Chronos is able to exploit longitudinal CRISPR data for improved inference. Additionally, it accounts for multiple sources of bias and can effectively share information across screens when jointly analyzing large datasets such as Project Achilles and Score. We show that Chronos outperforms competing methods across a range of performance metrics in multiple types of experiments.

Introduction

Genome-wide and large sub-genome loss of function CRISPR screens are increasingly important tools for understanding gene function in both normal and disease states. In a typical

experiment, cells are infected with a library of single-guide RNAs (sgRNAs) targeting genes of interest. The CRISPR-Cas9 system is less prone to the widespread off-target effects that occur in RNAi experiments(Ma et al., 2006). However, a number of other artifacts have been observed in pooled CRISPR screens which can complicate our ability to identify the true effect of gene knockout on cell fitness. These challenges include how to: interpret discrepant data for sgRNAs targeting the same gene, including identifying and correcting for variable sgRNA efficacy(Hsu et al., 2013); correct for nonspecific CRISPR-cutting induced toxicity, which causes a gene-independent depletion of sgRNAs targeting amplified regions(Aguirre et al., n.d.); reduce bias when comparing screens due to variable screen quality(Dempster et al., 2019); and address incomplete phenotypic penetrance due to heterogeneity in double-stranded break repair outcomes(Michlits et al., 2020).

A number of methods have been developed to address various combinations of these concerns. To combine sgRNA results into gene scores in a more robust manner than naive averaging, RIGER(Luo et al., 2008), RSA(König et al., 2007), and STARS(Doench et al., 2016) use statistical tests of guide rank significance to generate gene scores, while screenBEAM uses a Bayesian hierarchical model where variation across reagents are modeled as random effects(Yu et al., 2016). The Bayesian Analysis of Gene Essentiality (BAGEL) algorithm makes use of known essential and nonessential genes to estimate the probability that the sgRNAs targeting a given gene represent a true dependency(Hart & Moffat, 2016).

A well-known cause of variation in CRISPR systems is the variable on-target efficacy of individual sgRNAs. Given multiple screens with the same library, one can attempt a more sophisticated approach where the efficacy of different sgRNAs is inferred directly from the data

and thereby estimate gene fitness effects with greater weight placed on the more efficacious sgRNAs. This approach forms the basis of MAGeCK-MLE(Li et al., 2015), CERES(Meyers et al., 2017), and JACKS(Allen et al., 2019). These approaches are similar in that they treat the log of the relative change in sgRNA abundance during the experiment (log fold change) as the product of sgRNA efficacy and true gene fitness effect, although they employ different statistical assumptions and methods.

To remove bias related to DNA cutting toxicity, CERES uses a nonlinear model to estimate the fitness effects resulting from multiple DNA cuts and infers this relationship for each cell line using its measured copy number profile as input (Meyers et al., 2017). CRISPRCleanR (CCR) is an unsupervised approach for genome-wide screens: reagents are arranged according to their targeted location on the genome, and regions of systematic guide enrichment or depletion are shifted to the global mean on the assumption that they reflect a copy number alteration(Iorio et al., 2018). CCR is a “pre-hoc” method that should be used prior to the inference of gene fitness effects. Weck et al. introduced a pair of pre-hoc methods for copy number correction(De Weck et al., 2018), with a local drop out method similar to CCR and a generative additive model method similar to the CERES model of CN effect. Wu et al. introduced an update to MAGeCK-VISPR that similarly uses a paired approach: linear regression with a saturation value for cell lines that have CN profiles and a CCR-style alignment for cell lines that do not have CN profiles.(Wu et al., n.d.)

For analyses that compare gene essentiality estimates across screens, variation in screen quality can lead to significant biases(Boyle et al., 2018; Dempster et al., 2019; McFarland et al., 2018). To address this, Boyle et al. introduced a method for identifying and removing principal

components that reflect screen quality biases based on observed gene fitness effects of known non-essential genes(Boyle et al., 2018). A related approach is used to remove screen-quality-related principal components in the CERES data for the Cancer Dependency Map (DepMap) project(Dempster et al., 2019).

Although these methods address individual confounders in CRISPR screens, no existing method addresses all of them. Additionally, CRISPR screens are confounded by incomplete penetrance of the gene knockout phenotype. Given a single late time point measurement, poor knockout of a highly essential gene and complete knockout of a weakly essential gene may result in equivalent depletions, although the true phenotype is very different(Michlits et al., 2017). This ambiguity can be resolved by measuring fitness at multiple late time points, and with the falling costs of sequencing, an increasing number of experiments do so(Marcotte et al., 2016; Tzelepis et al., 2016). Marcotte *et al*. developed siMEM for combining multiple time points in the context of RNAi experiments(Marcotte et al., 2016). However, siMEM assumes sgRNA dropout occurs exponentially in time. This is a poor fit for CRISPR screens in which some clones escape gene knockout completely. For sgRNAs targeting essential genes, clones with intact function will eventually account for almost all reads and dropout will saturate at a finite value.

To simultaneously address these known challenges, we developed Chronos, an explicit model of cell population dynamics in CRISPR knockout screens. Chronos addresses sgRNA efficacy, variable screen quality and cell growth rate, and heterogeneous DNA cutting outcomes through a mechanistic model of the experiment. Like MAGeCK, but different from other approaches, Chronos also directly models the readcount level data using a more rigorous negative binomial

noise model (Anders & Huber, 2010), rather than modeling log-fold change values with a Gaussian distribution as is typically done. Copy number related biases are removed using an improved model that accounts for multiple sources of bias. We find that Chronos outperforms other methods on most benchmarks with a single late time point, and improves performance considerably when used to model data with multiple time points.

Results

Model

Chronos is a mechanistic model designed to leverage the detailed behavior of pooled CRISPR experiments to improve inference of gene essentiality. It models the observed sgRNA depletions across screens and time points to determine the effect of gene knockout on cell growth rate, along with other parameters. CRISPR knockout (KO) of a gene is an inherently stochastic process. Some sgRNAs will fail to cut their target(Tálas et al., 2021). In the event that they succeed and double-stranded break repair results in an insertion or deletion, in-frame mutations occur in about 20% of cases and may leave protein function intact(Michlits et al., 2020). Thus the result of introducing an individual sgRNA reagent in a population of Cas9 positive cells is heterogeneous, with outcomes including total loss, heterozygous loss, partial loss of function mutations, or completely conserved function(Shi et al., 2015). The Chronos model simplifies this range of outcomes to a binary pair of possibilities: total loss of function or no loss of function (**Fig. 1a**). Cells in the latter group will continue proliferating at the original, unperturbed rate. Those in the former will proliferate at some new rate reflecting the effects of a given gene perturbation on cell growth, which is typically the desired readout from the experiment. Concretely, for an sgRNA *i* targeting gene *g* in cell line *c*, we model the number of cells N_{cj} with the sgRNA at time *t* after infection as

$$N_{cj}(t) = N_{cj}(0)(p_{cj}e^{R_{cg}{}^{*}t} + (1 - p_{cj})\, e^{R_c\, t})$$

where *t*=0 is the time of infection, p_{cj} is the probability that the sgRNA *j* achieves knockout of its target in cell line *c*, R_c is the unperturbed growth rate of the cell line, and $R_{cg}{}^{*}$ is the new growth rate caused by knockout of the targeted gene in the given cell line. For Chronos, we define gene fitness effect as the fractional change in growth rate $r_{cg} = R_{cg}{}^{*}/R_c - 1$. Gene fitness effects are the primary desired output for this type of experiment.

A wide range of efficacy for sgRNAs in abrogating protein function has been reported(Doench et al., 2016). Additionally, we have observed in Project Achilles that screen quality (determined by separation of positive and negative control gene fitness effects) varies substantially across cell lines, due to variable Cas9 activity or other factors(Dempster et al., 2019). We therefore approximate the knockout probability per sgRNA and cell line as the product of a per-line and per-sgRNA factor, both constrained to the interval [0, 1]: $p_{ci} = p_c\, p_i$. The model also includes a gene-specific delay d_g between infection and the emergence of the knockout phenotype, measured in days. Finally, pooled screen sequencing data does not measure the number of infected cells N_{ci} directly but only the proportion of all reads that map to a particular sgRNA. This we assume has an expected value equal to the proportion of cells with that sgRNA : $\langle n_{ci} \rangle = N_{ci} / \sum_i N_{ci}$. Let the Chronos estimation of $\langle n_{ci} \rangle$ be ν_{ci}. Then,

$$\nu_{ci}(t) = Z_{ci}(t) / \sum_i Z_{ci}(t)$$

where

$$Z_{ci}(t) = \nu_{ci}(0)\,(1 + p_c p_i (e^{R_c r_{cg}(t - d_g)} - 1)) \quad \forall \quad t \geq d_g$$

$$Z_{ci}(t) = \nu_{ci}(0) \quad \forall \quad t < d_g$$

The parameters are estimated so as to maximize the likelihood of the observed readcounts under a negative binomial distribution, as detailed in the Methods section. In addition to inferring gene fitness effects, Chronos includes tools to remove suspected clonal outgrowth(Michlits et al., 2017) from read count data and to remove copy number (CN) biases from the inferred gene fitness effect (as described further below). A typical workflow proceeds as shown in **Fig 1b**. Note that the efficacy terms p_c and p_i in Chronos directly represent the probability of achieving functional knockout for a given guide and cell line, rather than enacting a simple multiplicative scaling of gene scores as in existing models. Additionally, the parameters p_c and R_c together account for much of the variation in screen quality which is a major source of confounding in genomics screens(Boyle et al., 2018; Dempster et al., 2019; McFarland et al., 2018). As a result, Chronos does far better at aligning good and poor-quality screens than competitors (**Supplementary Fig. 1a**): without scaling, the standard deviation of the median of essential gene effects within cell lines in Chronos is 58%-79% lower than its competitors. Any algorithm can improve its alignment by scaling the data after fitting, for example forcing the median of common essential gene scores to be -1 in each cell line(Dempster et al., 2019); however, one cost of doing so is an expansion in the noise of the lowest-quality screens (**Supplementary Fig. 1b**). Further, scaling does not eliminate biases related to variable screen quality, and in some cases can exacerbate these biases (**Supplementary Fig. 1c**).

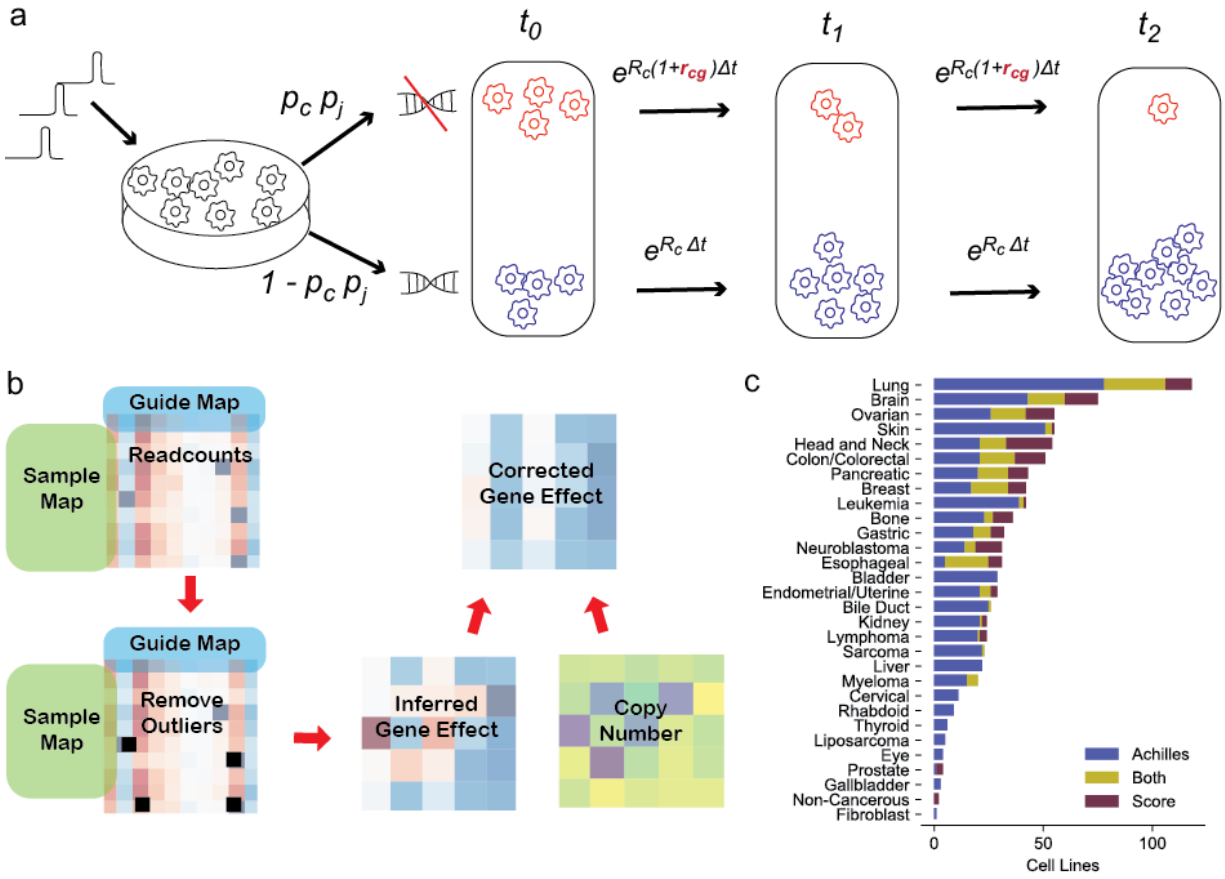
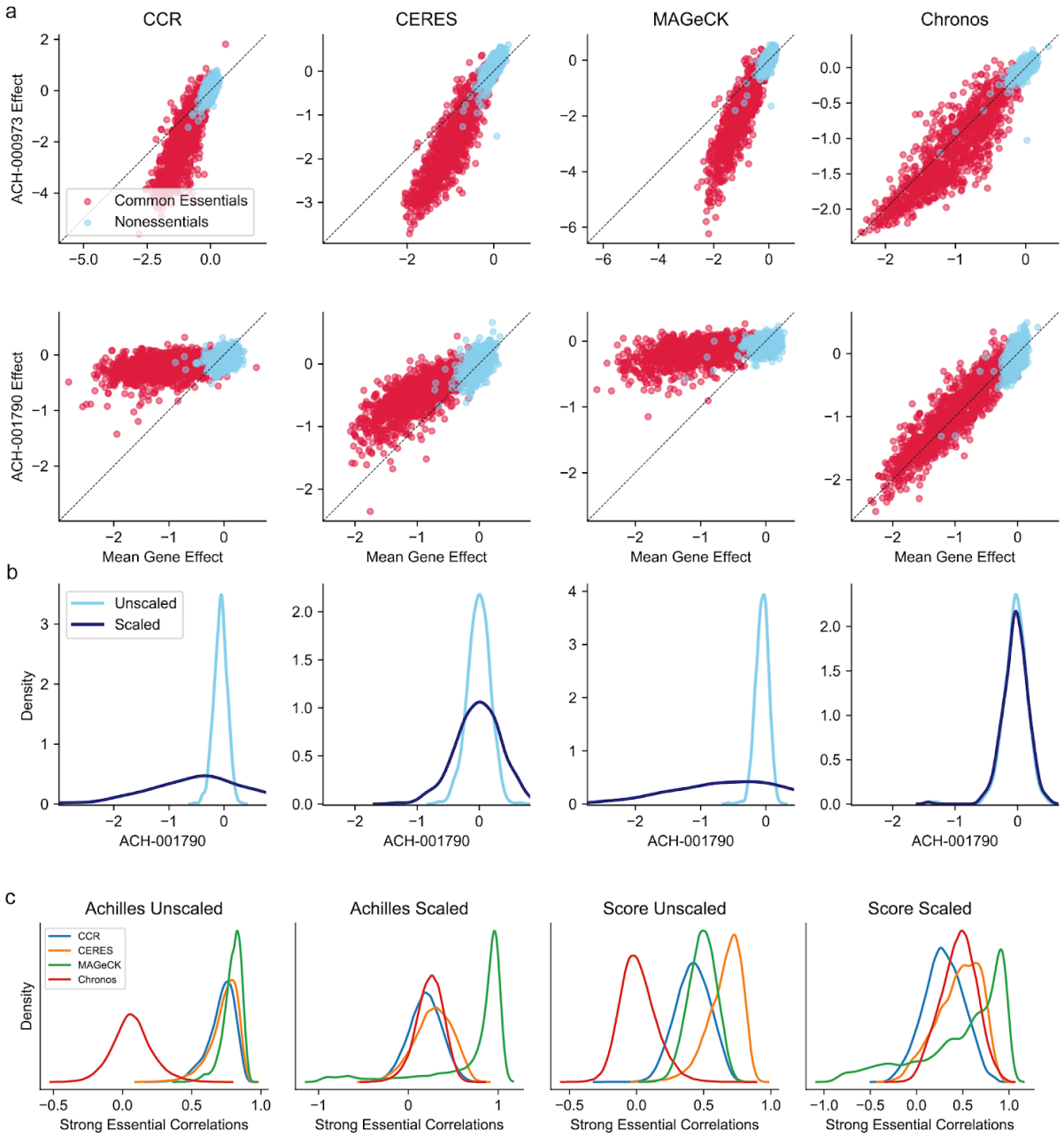


Figure 1: Overview of Chronos. **a.** Illustration of the model for the simplified case where an sgRNA *j* is introduced targeting essential gene *g* in cell line *c*. Stochastic double-strand repair divides the population of infected cells into two groups, one with (top) and one without (bottom) successful gene knockout, which proliferate at different rates. Successful knockout probability is the product of a per-sgRNA probability p_i and a per-cell line probability p_c. The population of cells measured at each subsequent time point is a mix of these two populations. Chronos infers the relative change in growth rate r_{cg}. **b**. Typical workflow of a Chronos run. Readcounts driven by outgrowing clones are removed, then gene fitness effects inferred using the readcount matrix, sequence map, and guide map. The inferred gene fitness effects are then corrected for copy number effects. **c**. Number of cell lines in each lineage for the Achilles and Project Score datasets used for comparison.



Supplementary Figure 1: Screen quality bias. **a**. Unscaled gene fitness effects for common essential and nonessential genes in an Achilles screen with good quality (top) and poor quality (bottom), plotted against the average gene fitness effects for those genes across cell lines. Each point is a gene. **b.** A simple strategy to reduce screen quality bias is to scale each screen such that the median of common essentials is the same value (-1 here). However, this can dramatically expand the noise in low quality screens. The two distributions show the width of the nonessential gene fitness effects in ACH-001790 (RH18-DM) before and after scaling by common essentials. **c.** Mutual correlations between gene effects for all possible pairings of the 300 strongest common essentials (most negative mean gene effect) for each version of the data, before and after scaling. Without scaling, Chronos has the lowest mean correlation, while other methods show indiscriminate correlation between all pairs of strong common essentials. After scaling, all methods exhibit indiscriminate correlation.

Comparison

We first evaluated Chronos on the two largest public datasets of human genome-wide CRISPR screens: projects Achilles(DepMap, 2020a) and Score(Behan et al., 2019) containing data from 769 and 317 screens, respectively (**Fig. 1c**). Among the many possible algorithms and combinations of algorithms, we selected three for comparison to represent the space of available choices. We chose CCR as a minimally-processed version of the data which illustrates only the effect of unsupervised copy-number correction(Iorio et al., 2018). We chose MAGeCK-MLE (MAGeCK) to illustrate an algorithm which similarly estimates guide efficacy and corrects for copy number while using a more statistically sound negative binomial model for screen noise(Li et al., 2015). Finally, we chose CERES as a method that combines guide efficacy estimates and copy-number correction with information sharing across cell lines via a hierarchical prior on the gene fitness effects(Meyers et al., 2017). CCR and CERES are additionally noteworthy as the primary processing methods currently used for CRISPR data from the Sanger and Broad DepMap projects respectively(Behan et al., 2019; DepMap, 2020a), while MAGeCK is a common choice for analyzing smaller screens. To focus on the results produced directly from the algorithms themselves, we did not perform any of the batch corrections described in Dempster et. al.(Dempster et al., 2019) or Pacini et al.(Pacini et al., 2020) However, for most results we did normalize the CERES and CRISPRCleanR data so the median of common essential gene fitness effects in each cell line is -1, as described in Meyers et. al.(Meyers et al., 2017). Some MAGeCK-processed cell lines had control separation too low for this to be a reasonable strategy (**Supplementary Fig. 1b**).

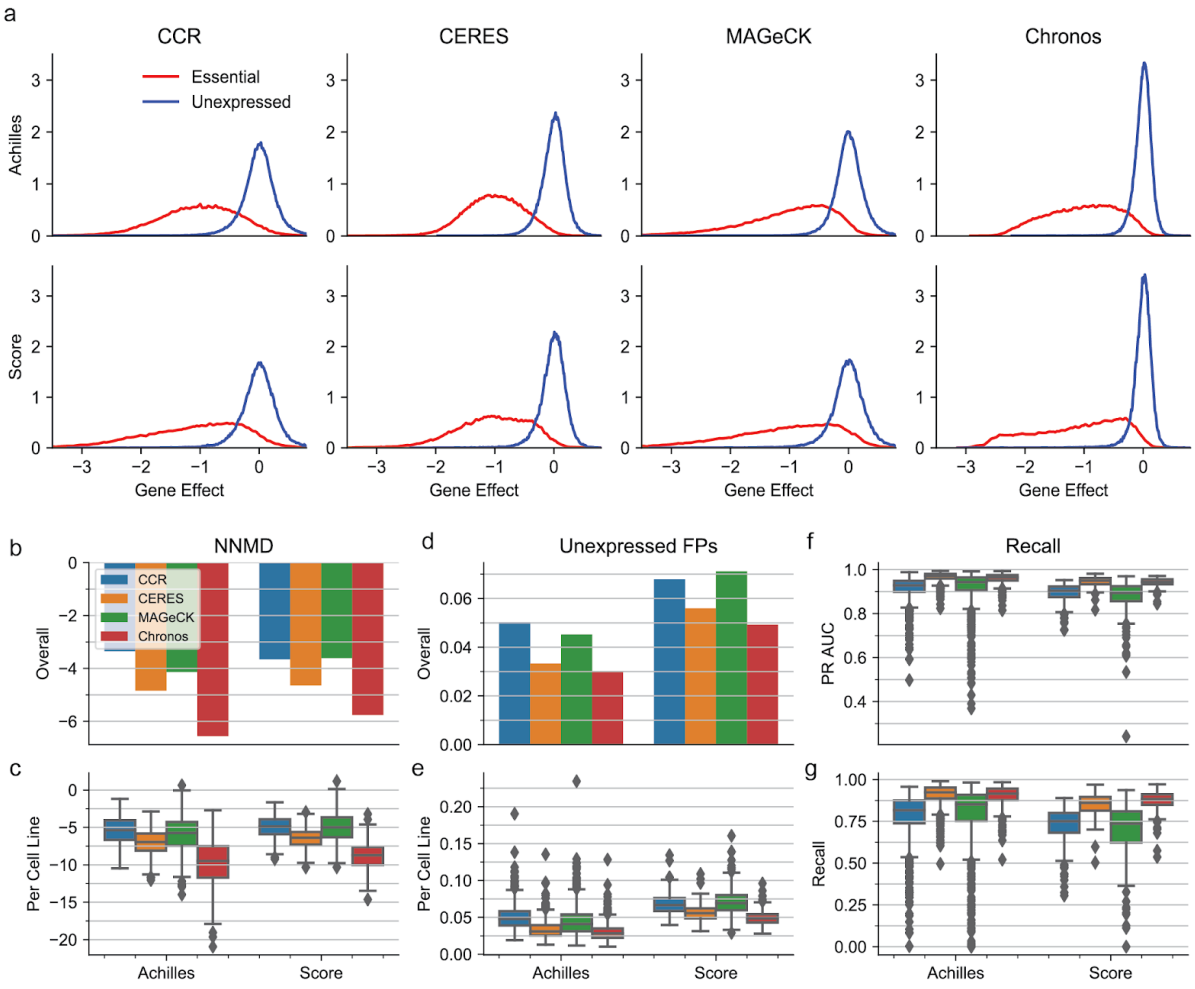


Fig. 2: Comparison of positive-negative control gene separation across methods for Achilles and Score datasets. **a**. The distribution of all gene fitness effects for unexpressed (negative control) genes and common essential (positive control) genes. Unexpressed genes are identified individually for each cell line (Methods). **b.** Global separation (pooled across cell lines and genes) between gene scores for common essential genes and unexpressed genes in the cell line where they are unexpressed. Separation computed using null-normalized median difference (NNMD). More negative values indicate stronger separation. **c**. NNMD for individual cell lines. Boxes indicate the interquartile range (IQR) of NNMDs across oncogenes. Whiskers extend to the last point falling within 1.5 x IQR of the box, and lines indicate medians. **d**. Estimated false positive rate, based on the total percentage of unexpressed genes scoring in most depleted 15% of gene scores within cell lines. **e.** Fraction of unexpressed genes scoring as false positives in individual cell lines. **f.** Area under the precision/recall curve (PR AUC), where recall is the number of common essential gene scores that can be recovered at a given precision. **g.** Fraction of possible common essential gene hits identified at 90% precision in individual cell lines.

The most straightforward indicator of success for a method is the separation it achieves between control sets of known essential and non-essential genes. We used a predefined set of common essential genes, which were identified from independent data, as positive controls(DepMap, 2020a). For each cell line, we used genes that are not expressed in that line as negative controls. The overall distributions of positive and negative control gene fitness effects are shown in **Fig. 2a**. We measured control separation in four ways, in increasing order of abstraction. First, we computed null-normalized median difference (NNMD) for all gene fitness effect scores (**Fig. 2b**) and for each individual cell line (**Fig 2c**). It reports the difference between the medians of the control sets normalized by the median absolute deviation of the negative controls. Alone of the measures used here, NNMD can be altered by rank-preserving transformations of gene scores. Chronos outperformed CERES by 40% (ratio of cell line means) in Achilles (paired Student's t-test $p = 7.2 \times 10^{-276}$, $N = 767$) and 36% in Project Score ($p = 9.5 \times 10^{-108}$, $N = 317$), which in turn outperformed the other two methods.

We next considered an estimate of how many false positive gene fitness effects would be identified by each method. We counted the total number of instances of unexpressed genes scoring in the 15% most depleted gene scores of a cell line (**Fig. 2d,e**), where 15% was chosen because previous estimates indicate that about 15% of genes are true dependencies in a given cell line(Dempster et al., 2019; DepMap, 2020b). Chronos outperformed CERES by 9.8% in Project Achilles ($p = 2.0 \times 10^{-50}$) and 12% in Project Score ($p = 6.6 \times 10^{-28}$).

We then considered the number of known common essential genes that can be recalled as hits with a given precision in each cell line. We measured the total area under the recall/precision

curve (PR AUC) for identifying all common essential gene scores (**Fig. 2f**) and the number of common essentials that can be recovered in each cell line at 90% precision (**Fig. 2g**). In all cases Chronos and CERES outperformed CCR and MAGeCK. CERES demonstrated a slight but significant improvement of 0.33%-0.63% by PR AUC in both datasets and by recall in Achilles (largest $p = 6.1 \times 10^{-4}$, $N = 241$). However, Chronos recalled on average 2.8% more essentials than CERES in Project Score ($p = 4.2 \times 10^{-14}$, $N = 241$).

Improvement with Multiple Time Points

The preceding analyses are performed entirely on datasets for which there is only one late time point measured per library. We next considered what advantages Chronos could offer when able to exploit sgRNA abundance measured across multiple time points. For this analysis, we used the study by Tzelepis *et al*.(Tzelepis et al., 2016) which contains CRISPR KO results for the cell line HT-29 measured at days 7, 10, 13, 16, 19, 22, and 25 post-infection. With only one cell line, we are unable to run CERES. Again we used the metrics NNMD, unexpressed false positives, and PR AUC, to assess positive versus negative control separation by Chronos when supplied differing numbers of time points in all possible combinations. **Fig. 3** illustrates the effect of adding additional time points on each of these three metrics of control separation. All three metrics showed improvement with increasing numbers of time points for all methods, but Chronos showed the greatest improvement. To verify that the Chronos population dynamics model effectively utilizes the data from multiple time points, beyond the simple denoising effect that might be expected from combining biological replicates, we compared the benefit of providing multiple time points simultaneously versys running Chronos separately for each of the same time points individually and taking the median of the results. For all metrics, Chronos performed better when provided multiple time points, and with three or more time points

outperformed all competitors on all metrics. We conclude that the Chronos dynamical model is able to exploit time-series data for improved performance beyond what could be achieved by naive averaging.

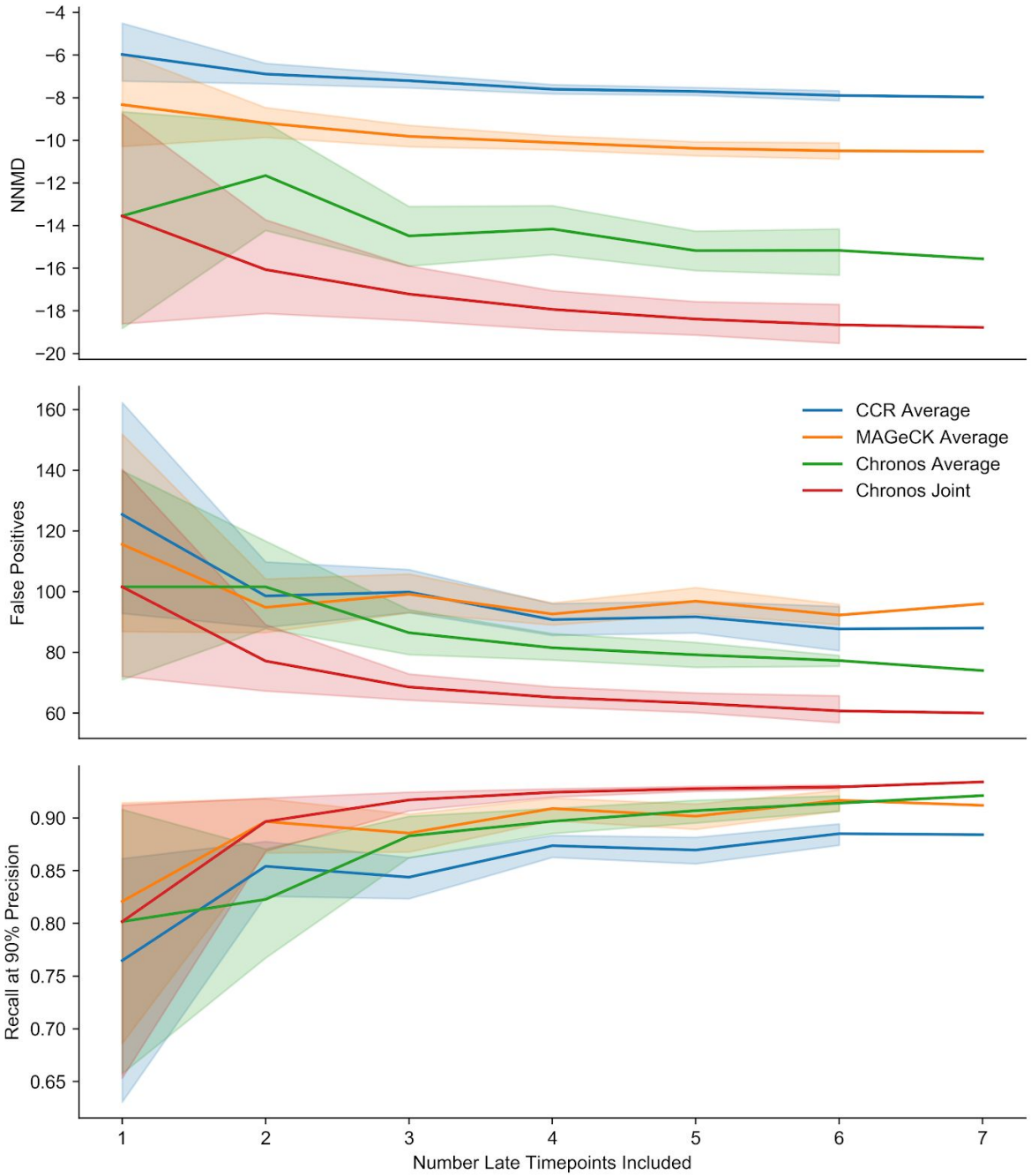


Fig. 3: Performance improvement with additional time points. The shaded area shows the 95% confidence interval for the (7 choose *n*) possible permutations of *n* measured time points that can be supplied to an algorithm. “Average” results are the performance achieved by taking a subset of the seven individual runs with a single late time point with a given algorithm and taking the median of their gene fitness effects. Joint results are obtained by running Chronos with a subset of multiple late time points simultaneously.

Performance in Individual Screens

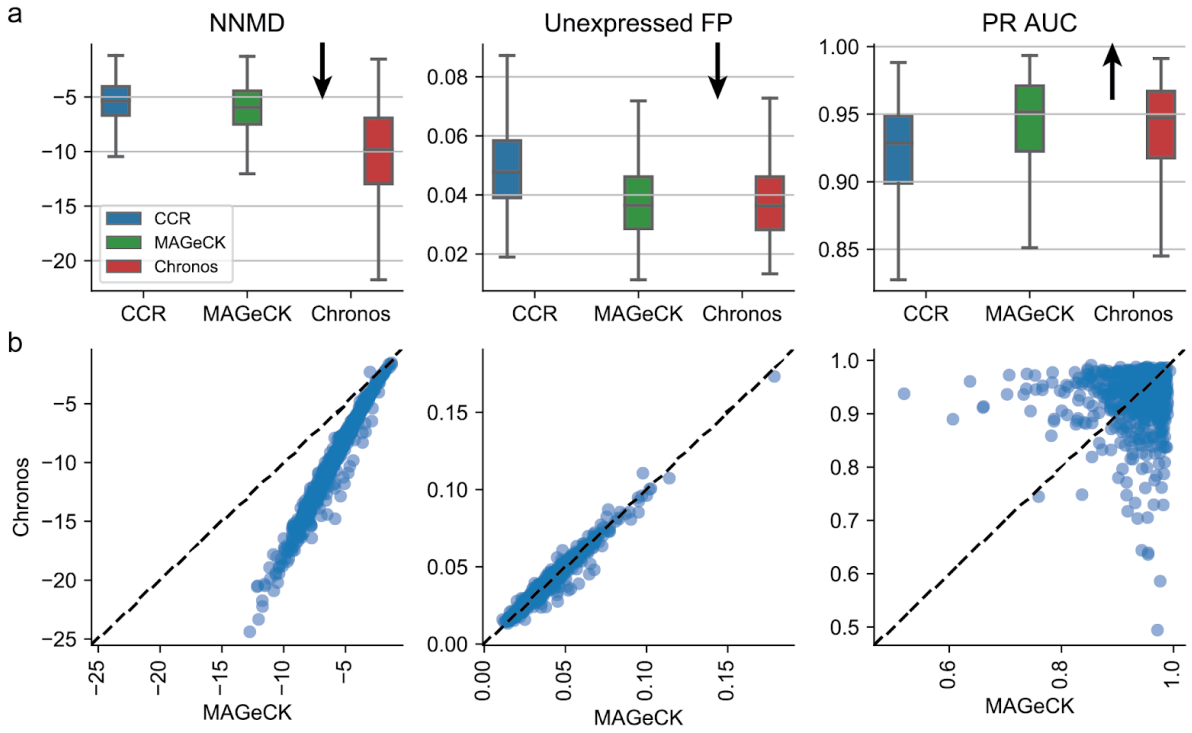


Fig. 4: Performance of Chronos on individual screens. Control separation for runs performed on Achilles screens one cell line at a time, evaluated using the metrics described in Fig. 2. **a**. The distribution of scores across cell lines. For CCR, this is equivalent to the earlier results since it analyzes each screen independently. Arrows indicate the direction of increasing performance for the metric. **b.** Chronos vs. MAGeCK performance, where each point is a cell line.

Many of the most powerful features in Chronos depend on sharing information across cell lines or time points. However, unlike CERES, Chronos can also be applied to analyze individual screens with a single late time point and can still benefit from properly accounting for read count noise. To assess its performance in this scenario, we ran Chronos and MAGeCK on each

Achilles cell line individually and quantified separation of positive and negative control genes as above (**Fig. 4**). Measured by NNMD, Chronos outperformed MAGeCK by 67% (related *t*-test $p = 4.8 \times 10^{-260}$). When using the two rank-based metrics Chronos and MAGeCK performed very similarly. The discrepancy between different metrics indicates that the ordering of genes is similar in Chronos and MAGeCK, but Chronos produces a tighter negative control distribution as shown in **Fig. 2a**. We note that Chronos and MAGeCK had high agreement with respect to which lines had the best quality as assessed by either NNMD or unexpressed false positives (Pearson's R = 0.99 and 0.98, *p* less than precision). However, PR AUC measured screen quality had no apparent agreement between the methods (R = 0.02, $p = 0.52$). This finding indicates that PR AUC may be a less stable metric than others for evaluating control separation. Overall, Chronos appears to perform favorably compared to existing methods, even when unable to integrate data across multiple late time points or multiple screens.

Identification of Selective Dependencies

The above analyses largely focus on the ability of each model to differentiate essential and non-essential genes for a given cell line, using a shared set of common-essential genes as positive controls. However, large-scale datasets, such as Project Achilles, are particularly powerful for identifying selective dependencies – genes that have a fitness effect in only a subset of the cell lines – which may represent cancer-selective vulnerabilities. Identifying selective dependencies across screens and cell lines presents a distinct challenge versus recovering known common essential genes. We thus evaluated the performance of algorithms in this context.

To benchmark estimation of selective dependencies, we first analyzed OncoKB genes that exhibit gain- or change-of-function alterations annotated as oncogenic or likely oncogenic(Chakravarty et al., 2017). After filtering as described in Methods, we retained 40 oncogenes known to induce oncogene addiction with activating alterations that had at least one annotated alteration in at least one cell line in at least one dataset. These gene scores were considered selective positive controls in cell lines which had any annotated (likely) gain- or switch-of-function alterations, and selective negative controls in all other lines. There were a median of four cell lines with annotated alterations for each oncogene in the Achilles dataset and two cell lines in Project Score.

To determine how well gene fitness effect profiles for oncogenes were stratified by the known alterations, we measured the difference between cell lines with and without an indicated gain- or switch-of-function alteration using NNMD in each dataset (**Fig. 4a**). No significant differences were observed between pair of methods due to the small number of oncogenes and small number of positive examples in each gene (smallest related *t*-test p=0.095 between CERES and MAGeCK in Achilles, N=40). To address this problem, we pooled all selective controls to increase statistical power (**Fig. 4b**). Chronos outperformed its closest competitor (MAGeCK) in both datasets by 8.09% in Achilles (permutation testing $p < 1 \times 10^{-3}$, N = 1000). The difference of 3.16% in Project Score was not significant (p = 0.31). Measuring separation by PR AUC, Chronos outperformed MAGeCK by 13.5% in Achilles data ($p < 1 \times 10^{-3}$), and by 12.1% in Project Score (p = 0.022, **Fig. 4cd**).

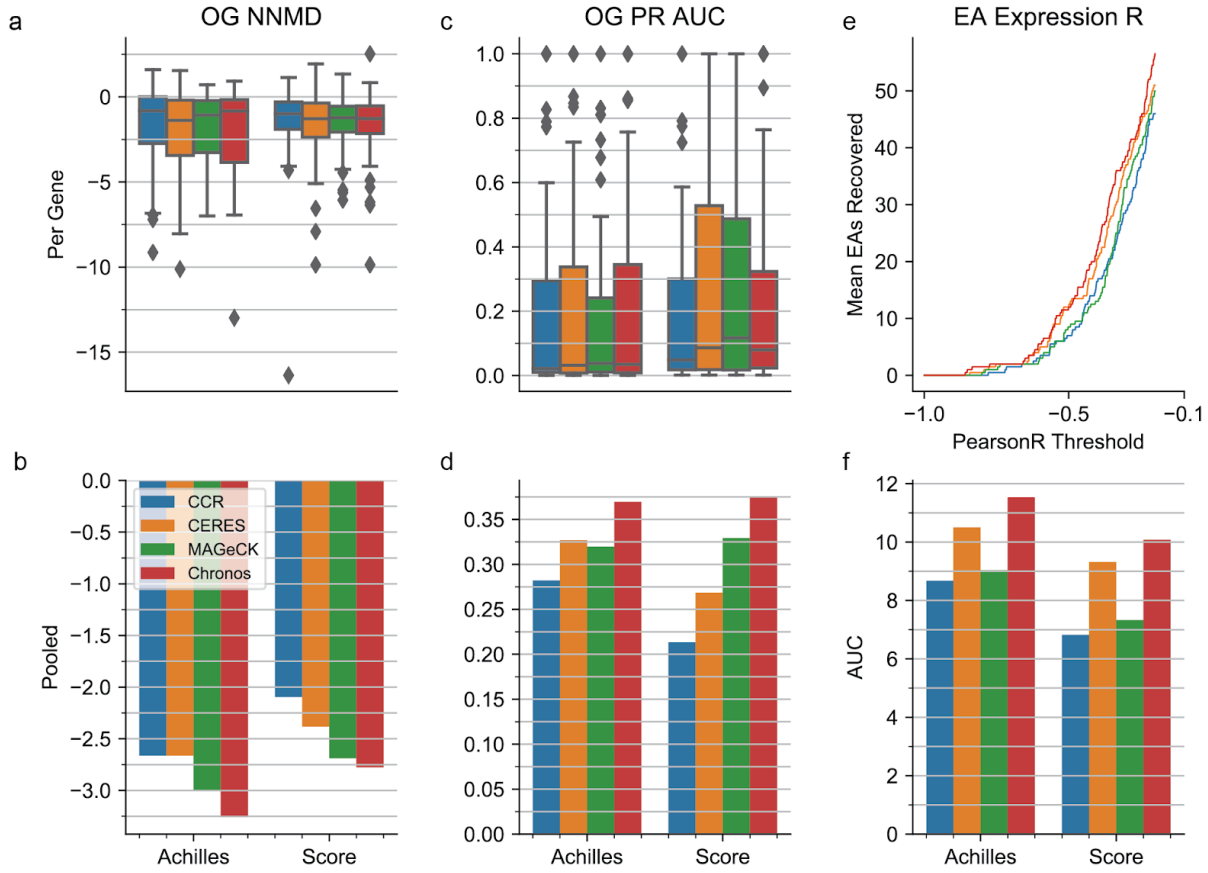


Fig. 5: **a**. Example of the influence of the CERES hierarchical regularization on gene fitness effect in the unscaled CCR Achilles screen of cell line ACH-000235/PANC0403. Each point is a gene, with the y axis showing its fitness effect in the specific cell line and the x axis showing the effect across cell lines. For clarity only 1% of gene scores are shown. Arrows indicate the direction and magnitude of the regularization penalty. PANC0403 has activating mutation G12D in KRAS (red). In the limit of extreme regularization, fitness effects for each gene would be moved to the corresponding arrow tips. **b**. Similar for the modified hierarchical penalty used by Chronos with a Gaussian kernel. **c**. The distribution of NNMDs for each oncogene between positive and negative cell lines. **d.** NNMD of selective positive controls from selective negative controls across all retained oncogenes. **e**. PR AUC for separating cell lines with an indicated alteration from those without for individual oncogenes. **f.** The same for separating all selective positive controls from selective negative controls. For the indicated threshold, the number of expression addictions whose gene fitness effects are more negatively correlated with their expression than the threshold, averaged between the Achilles and Project Score datasets. **j**. The area under the curves as in (i) for the individual datasets.

To evaluate the ability of models to identify expression-based gene dependency profiles, we used the expression addictions previously identified in Achilles RNAi data (Tsherniak et al.,

2017). These genes exhibited selective essentiality in cell lines in which they were highly expressed. After subsetting to those present in all algorithms and datasets, and removing any common essential genes (identified from the DepMap releases (DepMap, 2019, 2020a)), we had 106 putative expression addiction genes. We compared the number of potential expression addictions whose gene fitness effects were negatively correlated with their own expression below a given correlation threshold between datasets, varying the threshold from -1 to -0.1 (**Fig 4e**). Chronos consistently identified more expression addictions as correlated with their own expression at a given threshold than its nearest competitor CERES, resulting in a significantly greater AUC for Achilles (8.95% improvement, two-tailed permutation p = 0.047, N = 40,000). Compared to CERES in Project Score, Chronos showed a similar improvement (7.56%) but the difference was not statistically significant due to greater variability in correlation recall (**Fig 4f**).

Accounting for Copy-number Biases

It is important for any model for inferring gene KO effects from CRISPR screens to account for the gene-independent DNA-cutting toxicity effect(Aguirre et al., n.d.). A precise description of the causes of this cutting toxicity is still lacking, and current literature suggests it may be complex. For example, Gonçalves et al. argue that the copy number effect arises in tandem repeats alone(Goncalves et al., 2018). Furthermore, the relationship between gene fitness effect and copy number varies for different types of genes. For most genes, the observed sgRNA abundance is negatively correlated with copy number across cell lines, but the opposite is true for essential genes (**Fig. 6a**), likely owing to the decreased probability of achieving complete loss-of-function when more copies of a gene are present. In fact, for essential genes, the average relationship of gene scores with copy number is non-monotonic (**Fig. 6b**), further highlighting the complexity of the effect. Instead of developing a mechanistic model for the copy

number effect, we created a new post-hoc method for correcting copy-number-related biases that can be applied to Chronos gene fitness effects or to analogous output from other modeling approaches. As the copy number effect is a function of both the mean fitness effect of a gene and its copy number in a given cell line, our correction uses a 2D spline model to capture and remove the systematic nonlinear dependence of gene scores on both parameters. Chronos first fits the spline model so that as much of the variance in the gene fitness effects as possible is explained by copy number effect, modulated by the mean gene effect (see Methods). The residues of the spline model are taken as the corrected gene fitness effects. Where CERES successfully corrects the cutting toxicity effect, it worsens the bias due to incomplete gene KO in common essential genes (**Fig. 6c**). In contrast, this new correction method incorporated with Chronos corrects both effects simultaneously. The Chronos copy number correction can be run independently on any gene fitness effect matrix.

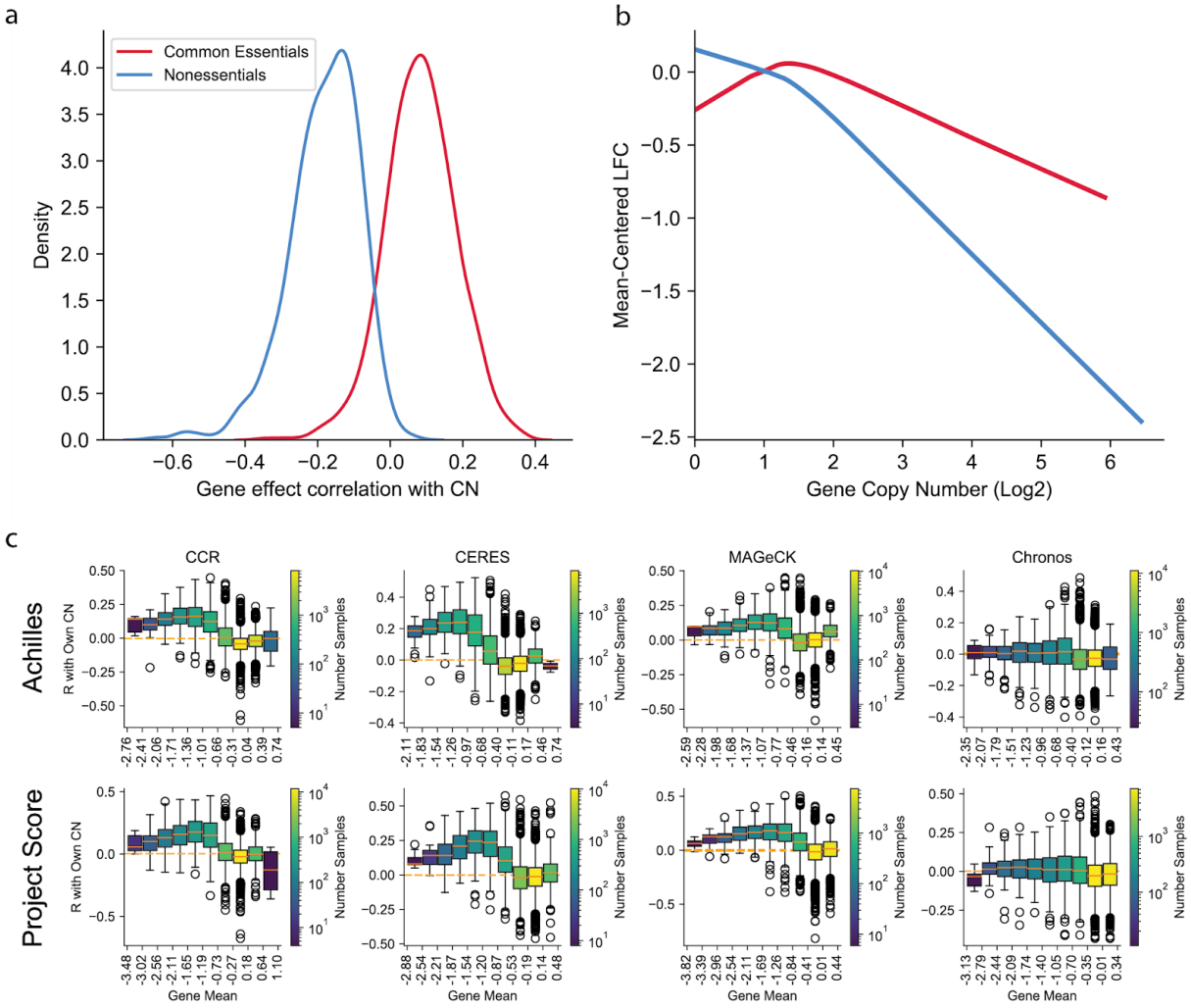


Fig. 6: **The copy number effects. a**. Distribution of correlations for uncorrected gene log fold changes with their own copy number across cell lines in Achilles for common essential and nonessential genes. **b.** Lowess smoothed trends for the mean-centered log fold change of known essential and known nonessential genes as a function of copy number. **c**. Per-gene correlations of gene fitness effects with its own copy number, binned by mean gene fitness effect. The boxes show the IQR for the correlations of genes in the given bin, whiskers extending to the last data point within 1.5 the IQR from the median.

Discussion

Here we have introduced Chronos, a new approach for estimating the fitness effects of gene knockout in CRISPR screens using a model of the cell population dynamics. We compared the performance of Chronos to existing methods in terms of separation of both global and cell-line-specific essential and nonessential genes. Across two independent large-scale CRISPR screening datasets, Chronos generally outperformed existing methods, with a few exceptions.

Chronos also outperformed alternative methods when applied to a single screen performed with many late time points. Importantly, performance for Chronos improved appreciably when incorporating longitudinal data into its dynamical model, compared with simply averaging together results based on different time points as a general denoising procedure. For example, with three late time points, mean performance for Chronos improved by 14% to 31% over its mean performance with only one time point. Chronos exhibited both better absolute performance and larger relative performance gains with additional time points than all competing methods, or than Chronos itself using an averaging strategy. This observation suggests that there may be significant additional value in measuring abundance across a series of time points as previously suggested(Marcotte et al., 2016), and that modeling cell population dynamics can increase the benefit of multiple time points.

In the more common context of a single late time point, a number of factors still contribute to improved performance for Chronos. These are 1) an improved strategy for correcting copy number related bias that produces less distortion in common essential genes, 2) a more

sophisticated regularization strategy that reduces bias in estimating selective dependencies while retaining effective information sharing and normalization across screens, 3) an improved generative model of the data at the readcount level, and 4) an explicit model of multiple "screen quality" factors that correct for systematic differences across screens. Additionally, unlike some competitors, Chronos does not assume either a genome-wide experiment(Iorio et al., 2018) or require multiple cell lines(Meyers et al., 2017). When analyzing a single screen with a single late time point, we found that Chronos performed as well or better than MAGeCK and CCR, probably due to its generative model. Thus it is a flexible method that provides improved estimates of gene essentiality from a variety of CRISPR experiments.

Methods

The Chronos Algorithm

Chronos maximizes the likelihood of the observed matrix of normalized readcounts (relative sgRNA abundance) with the following model:

$$\nu_{ci}^{L}(t > d_g^{L}) = \nu_{ci}^{L}(0)\,(1 + p_c^{L} p_i (e^{R_c^{L} r_{cg}(t - d_g)} - 1)) / Z_c^{L}(t)$$

where

- *c* indexes cell line, *i* indexes sgRNA, *g* indexes gene, *L* indexes library or batch, and *t* is the time elapsed since library transduction
- $\nu_{ci}^{L}(t)$ is the model estimate the normalized readcounts of sgRNA *i* in cell line *c* screened in batch or library *L* at time *t*
- $\nu_{ci}^{L}(0)$ is the model estimate of the normalized number of cells initially receiving

sgRNA *i*

- p_c^L and p_i are the estimated CRISPR knockout efficacies in cell line *c* with sgRNA *i*
- R_c^L is the estimated unperturbed growth rate of the cell line
- r_{cg} is the estimated relative change in growth rate for that cell line if gene *g* targeted by sgRNA *i* is complete excised
- d_g is the delay between infection and the onset of the growth phenotype
- $Z_c^L(t)$ is a normalization equal to the sum of the numerator over all sgRNAs *i* in the cell line for the given library and time point

Unless simultaneously fitting Chronos to data generated from multiple libraries, *L* can be ignored.

Previous models often treat sgRNAs targeting more than one gene as producing a linear addition of the gene fitness effects in log space. However, it is clear that such a treatment is not biologically supportable(Fortin et al., 2019), and the actual effect of simultaneous knockout is highly dependent on the pair of genes targeted(Najm et al., 2017). Consequently we do not attempt to model multitargeting sgRNAs in Chronos, and our implementation will raise an error stating so if they are encountered.

A naive approach to the distribution of readcounts is a multinomial likelihood. However, biological readcount data often have more dispersion than can be accounted for by multinomial or Poisson models(Anders & Huber, 2010; Li et al., 2015). The NB2 model is a frequently used parameterization of the negative binomial model that accepts an overdispersion parameter α. The likelihood of observing counts *N* when *μ* such counts are expected under the NB2 model is

$$\Pr(N|\mu,\alpha) = \frac{\Gamma(N+\alpha^{-1})}{\Gamma(N+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{1/\alpha}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{N}$$

We will treat α as a fixed hyperparameter and the observed readcounts as given, so only terms involving μ need to be retained. Thus the negative log likelihood becomes

$$\lambda = (N+\alpha^{-1})\ln(1+\alpha\mu) - N\ln\mu$$

Note that an equal change in the scales of N and μ is equivalent to a change in the scales of α and λ up to a constant. With an added constant to ensure the cost is zero when $N = \mu$ and some slight abuse of notation, the core cost function for Chronos reads

$$C = \sum_{L}\sum_{i\in L}\sum_{c\in L}\sum_{k\in c}\sum_{t\in L}(10^{6}\cdot {n_{kit}}^{L} + ({\alpha_c}^{L})^{-1})\ln\left(\frac{1+10^{6}\cdot{\alpha_c}^{L}\,{\nu_{ci}}^{L}(t)}{1+10^{6}\cdot{\alpha_c}^{L}{n_{kit}}^{L}}\right) - N\ln\left(\frac{{\nu_{ci}}^{L}(t)}{{n_{kit}}^{L}}\right)$$

where k is a replicate of a cell line screened and t both indicates the actual time elapsed since infection and indexes the readout at that time. Rather than estimating library size, we have simply fixed this cost to take in readcounts normalized to reads per million. We made this decision because there is not much evidence that the number of total readcounts found in a screen governs the statistics for the screen. Instead, users can supply the overdispersion parameter α either as a global hyperparameter or as an estimate per cell line and library using independent tools such as edgeR(Robinson et al., 2010). In our experience, using edgeR estimates of the overdispersion resulted in values so high for some cell lines that they effectively contributed nothing to the cost, despite having clear indications of signal. We opted instead to set a global value.

Constraints and Regularization

As written, the Chronos model is not identifiable. For example, one can multiply a value R_c by some value and divide all the corresponding values of r_{ci} by the same amount and produce exactly the same estimate. More subtle degeneracies exist between knockout efficacy and gene fitness effect and between gene fitness effect and the estimate of initial cell abundance $v_{ci}^{L}(0)$. Even for terms not exactly degenerate, the quality of the model fit is improved by regularization. We applied the following constraints and penalties:

$\nu_{ci}^{L}(0)$**:** This initial time point could in theory be inferred in the same way as any other parameter, with pDNA abundance supplied as a *t* = 0 time point. However, previous work has found evidence of some systematic errors in initial sgRNA abundance as measured in pDNA(Meyers et al., 2017) in which some sgRNAs appear to be consistently more or less abundant in cell lines than could be explained by either target knockout effect or the measured pDNA abundance. Additionally, most experiments batch pDNA measurements across many screens. Thus, we instead use the following:

$$\nu_{ci}^{L}(0) = n_{b(c)i\,0}^{L} \cdot \exp(\rho_{bi}^{L})$$

where $n_{b(c)i\,0}^{L}$ is the median measured relative abundance of sgRNA *i* in the pDNA batch *b* of cell line *c* in library *L* and ρ_{bi}^{L} is a parameter to be estimated. It is constrained to have mean 0 for the complete set of sgRNAs that target any specific gene (recall that Chronos does not accept sgRNAs annotated as targeting multiple genes). Additionally, it is subject to the penalty

$$C_\rho = \chi_\rho \frac{1}{K} \sum_{L} \sum_{b \in L} \sum_{i \in L} (\rho_{bi}^{L})^2$$

where we will use *K* here and elsewhere to indicate the total number of terms in the sums and χ_p is a regularization hyperparameter equal to 1.0 by default.

p_c^L: In the case of only one late time point, this cell line screen quality parameter is degenerate with both the unperturbed cell line growth rate R_c^L and (in a global sense) with the sgRNA quality parameter p_i^L. To avoid these degeneracies, we estimate this parameter directly from the fold change in the *n*th most depleted sgRNA at the last available time point for the library, where *n* is the 99th percentile of sgRNAs by default.

p_i: As we noted above, sgRNA efficacies in CERES tend to amplify the most extreme sgRNA results. Specifically, if three of four sgRNAs show little depletion in every line and one sgRNA shows substantial depletion in any cell lines, CERES will usually assign high efficacy to the outlying sgRNA and low efficacy to the others. To prevent Chronos from similarly chasing a single sgRNA, we force it to assign efficacy values near 1 to at least two guides with the following term:

$$C_p = \chi_p \frac{1}{K} \sum_g \sum_{i \in g} I_{gi} \, p_i^{-1}$$

where χ_p is a regularization hyperparameter set to 0.5 by default and I_{gi} is an indicator function with value 1 iff the sgRNA *i* is currently estimated to be the first or second most efficacious sgRNA for the gene *g*.

R_c^L: The per-cell line and library unperturbed growth rate is degenerate with the cell efficacy

and the individual rows of r_{cg} . We constrained it to have positive values and mean 1. Additionally, it is regularized with the penalty

$$C_R = \chi_R \frac{1}{K} \sum_L \sum_{c \in L} \ln R_c{}^L$$

where χ_R is a regularization hyperparameter with default value 0.01. The log penalty is chosen because it has little influence on a parameter constrained to have mean 1 unless some cell lines approach 0, a behavior we have observed occasionally with small internal datasets.

r_{cg} : The gene fitness effect matrix is regularized in two ways. First, the global mean is strongly regularized towards 0:

$$C_{r1} = \chi_{r1} \sum_c \sum_g r_{cg}$$

where χ_{r1} is a regularization hyperparameter with default value 0.1. The second regularization is a smoothed hierarchical kernel prior

$$C_{r2} = \frac{1}{K} \sum_c \sum_g \sum_h (\kappa_{g-h} (r_{ch} - \bar{r}_h)^2)$$

where κ_{g-h} is a kernel function of the rank distance of the means of the effects of genes *g* and *h*, and $\bar{r}_h$ is the mean value of gene *h* across cell lines. The kernel is a combination of two terms:

$$\kappa_{g-h} = \chi_h \delta_{gh} + \chi_k \frac{1}{b} \exp((g-h)^2 / 2\sigma^2)$$

where χ_h, χ_k and σ are hyperparameters with default values 0.1, 0.25, and 5, δ_{gh} is the Kronecker delta, and *b* is a normalization term that makes the sum of the Gaussian exponential 1. For computational efficiency, the support of κ is restricted to 3σ in either direction from zero.

For interpretability, we recommend users shift and scale the whole inferred gene matrix so the median of all nonessential gene scores is 0 and the median of all essential gene scores is -1 *globally*, not per cell line.

d_g^L: In theory, the gene knockout phenotype delay parameter could be inferred given sufficient time points, in particular time points measured relatively close to infection. In practice, we have never found a benefit to trying to do so. Instead, it is fixed as a hyperparameter with default value 3 days, which the authors have found roughly approximates the onset of a fitness phenotype for many essential gene knockouts.

Implementation and Fitting

Training Chronos consists of choosing the free parameters to minimize the combined cost function

$$C_T = C + C_\rho + C_p + C_R + C_{r1} + C_{r2}$$

We implemented the Chronos model in tensorflow v1.15 and use the native AdamOptimizer to train the parameters. The pDNA offsets ϱ are initialized to 0, the guide efficacies p_i are initialized to values near 1 with a small random negative offset, unperturbed growth rates R_c are initialized near 1 with a random gaussian offset (standard deviation 0.01), and for the gene fitness effect r_{cg} the gene means and the per-cell-line scores are separately initialized to very small random values in the interval [-10^{-4}, 0.5×10^{-4}].

We use a few tricks to improve learning performance. First, to ensure numerical stability in the exponents, time values are multiplied by a constant with default value 0.1, equivalent to measuring time in units of 10 days. Second, the core cost *C* is rescaled so it has an initial value of 0.67 by default, equivalent to renormalizing the hyperparameters. This ensures more consistent effects for the regularization terms in datasets with different sizes. With this adjustment, we have found that the default hyperparameters work well for all cases tested, including sub-genome libraries with small numbers of screens. Third, we find that using a fixed learning rate for AdamOptimizer causes either an initial explosion in the cost function before the optimizer begins minimizing, or else inefficient learning if the rate is very small. The final optimal parameters learned for both these cases are quite similar, but learning is inefficient. We therefore initialize the optimizer with a default maximum learning rate of 10^{-4}, rising exponentially over a burn in period of 50 epochs until it reaches a default value of 0.02. Finally, we optimize the gene fitness effect alone during the first 100 epochs. Because the core cost function is convex when gene fitness effect alone is considered, this helps ensure stability for the optimum found by Chronos. Training runs for a number of epochs specified by the user (default 801).

We have noticed a pattern of rare clonal outgrowth in CRISPR data, where a single sgRNA for a single replicate of a cell line will show unusually large readcounts. This may be due to a random fitness mutation or other artifact. We provide users an option to preprocess their readcount data by identifying and removing these rare events. Specifically, for cases where the log2 fold change of a guide in a biological replicate is greater than a specified threshold, and the difference between that log fold change and the next highest log fold change for an sgRNA targeting that gene in the given replicate is lower by a specified amount, the outgrower is NAed.

About 0.02% of all readcount entries in Achilles and Project Score were NAed for suspected clonal outgrowth.

Removing Copy Number Effect

To remove copy number bias, Chronos provides a function that accepts a matrix of gene-level copy number (processed as in DepMap: log2(*x*+1) where *x* is the relative copy number) and a matrix of gene fitness effect. It constructs a two-dimensional cubic spline representation for each gene fitness effect score in each cell line, where the first dimension is the copy number of the gene in a given line and the second is the mean effect of the gene across lines. By default, the model uses ten knots linearly spaced in copy number space and five knots in mean gene fitness effect space, spaced pseudo-exponentially (meaning, exponential if the first percentile mean gene fitness effect is taken as 1 and the other mean gene fitness effects shifted accordingly), so that knots are more concentrated in the region of strong negative mean gene fitness effect where the copy number effect changes fastest. The model for the copy number effect is then written as follows:

$$y_{cg} = w_c \sum_k \theta_k B_{cgk}$$

where B_{cgk} is the spline basis representation, θ_k are the spline coefficients, and w_c is a per-cell-line parameter in the interval (0, 1] that weights the strength of the copy number effect in that line. The model minimizes the cost function

$$C_{CN} = \sum_{c,g} (r_{cg} - y_{cg})^2 + \chi_w \sum_c \ln(w_c)^2$$

then returns the residuals, which can be treated as bias-corrected gene fitness effect estimates for downstream analysis.

Algorithm Runs

For all algorithms, we began with the readcount tables provided for Achilles(DepMap, 2020a) and Project Score(Behan et al., 2019), which have each undergone different QC pipelines.

To compute log fold change (LFC), we calculated the base-2 log of reads per million (RPM) + 1 and subtracted the pDNA values for the appropriate batch from the late time points. For Achilles data, which has multiple pDNA measurements, we summed pDNA measurements from the same batch prior to computing RPM. To calculate a naive gene fitness effect score for **Fig. 5ab**, we filtered out all sgRNAs with multiple gene targets and computed gene fitness effects per replicate as the median of the LFCs for all sgRNAs targeting the gene in question. We then computed cell line gene fitness effects by taking the median of replicate gene fitness effects.

All gene effect estimates were normalized (shifted and scaled) globally so that the median of the medians of reference nonessential genes was 0 and the median of the medians of essential genes was -1 across cell lines. This was done for visual interpretability only and has no effect on the metrics evaluated in this manuscript.

CCR was run per cell line in each dataset using the functions gwSortedFCs and correctedFCs from version 0.5 of the R package CRISPRCleanR. Log fold changes were first computed as described above. We excluded sgRNAs targeting more than one gene, and in the Avana library

those that had more than one exact match anywhere in the genome (hg38). Genes were further restricted to those with at least two remaining sgRNAs.

For Chronos, sgRNAs were filtered similarly. Suspected clonal outgrowths were removed with the methods described above and the model trained for 801 steps. The resulting gene viability effect matrix was then globally normalized as described above and then copy number corrected as described above using the DepMap gene level copy number data(DepMap, 2020a). Genes present in the gene viability matrix and not the copy number matrix were assumed to have normal ploidy (assigned value 1).

CERES results were taken directly from the Dependency Map file gene_effect_unscaled(DepMap, 2019, 2020a).

To run MAGeCK-MLE, we filtered the sgRNAs as described above. These cleaned tables were then split by pDNA batch. Due to issues with memory consumption (greater than 50GB) and run time (greater than 3 days) for larger batches, batches which consisted of more than 70 cell lines were further subdivided into sets of approximately 50 cell lines. Design matrices were constructed to indicate pDNA as the baseline and map experimental replicates to their respective cell line or time point. MAGeCK 0.5.9 was run with the “mle” subcommand in order to perform maximum-likelihood estimation of the gene essentiality (beta) scores. The sgRNA efficiency was iteratively updated during EM iteration using the “--update-efficiency” flag. In order to further reduce memory usage and runtime, permutation was performed on all genes together, rather than by sgRNA count, using the “--no-permutation-by-group” flag and genes with more than 6 sgRNAs, a total of 33 in the Achilles and 673 in Project Score, were excluded

from the maximum-likelihood estimation calculation using the "--max-sgrnapergene-permutation" parameter. The effects of copy number variation were corrected using the "--cnv-norm" parameter. The DepMap gene level copy number data (DepMap, 2020a) was converted to log2 and genes with missing data were imputed with zeros, indicating no change in copy number. The beta scores were then extracted from the outputted gene summary files and aggregated across batches to construct a gene fitness effect matrix for each dataset. Batch effects between the cell line batches were removed by ComBat(Johnson et al., 2007).

Runs for the single cell line analysis were performed in Chronos as described above for the full Achilles dataset, but filtering the read count matrix to include only one cell line at a time during inference. Gene effects from the runs for individual cell lines were stored and concatenated afterwards into a single matrix. A similar approach was used with MAGeCK. Neither Chronos nor MAGeCK single line runs were copy-number corrected.

Analysis Methods

Unless otherwise reported, all *p* values were calculated using the python scipy.stats package using the given test, and all analyses were restricted to genes in common across all versions of the data.

For each dataset, algorithm, and cell line, the NNMD score was calculated as the difference in the medians of positive controls and negative controls, normalized by the median absolute deviation of negative controls. Positive controls were taken as the intersection of common essential genes in the DepMap public 20Q2 dataset with the genes present in all datasets and algorithms. In each cell line, negative controls were unexpressed genes: those with expression

less than 0.01 in that cell line according to the DepMap public 20Q2 dataset. When such genes fell in the bottom 15% of all gene fitness effect scores for the cell line for a given algorithm and dataset, the score was considered an unexpressed false positive.

Precision and recall for global controls were estimated directly from the gene fitness effects of common essential and unexpressed genes in each cell line individually using the sklearn.metrics function precision_recall_curve.

To produce a list of selectively essential genes, we downloaded the OncoKB annotations for all variants (http://oncokb.org/api/v1/utils/allAnnotatedVariants, accessed July 15th, 2020)(Chakravarty et al., 2017). Initially 468 genes were present. Rows were filtered for alterations listed as “(Likely) Oncogenic”, with mutation effects in “(Likely) Gain of Function” or “(Likely) Switch of Function,” leaving 225 genes. We filtered out those not present in all versions of the Achilles and Project Score data, leaving 183 genes. We then removed those identified as common essential from the CERES gene effect scores in the DepMap releases for Project Achilles and Project Score, leaving 169 genes. DepMap fusion and mutation calls were used to identify cell lines with the indicated alteration, where the altered gene was treated as a selective positive control. For fusions, both gene members of the fusion were treated as selective positive controls. If no dataset had any cell lines with an indicated alteration for a gene, that gene was dropped, leaving 69 oncogenes. Finally, the 69 surviving oncogenes were manually curated to identify oncogenes known to induce oncogene addiction, leaving 40 genes. Reported metrics (NNMD and precision/recall curve) were calculated both per oncogene between cell lines with and without indicated alterations, and collectively after pooling all selective positive control and selective negative control gene fitness effects. To calculate permutation *p*-values for the pooled

case, we randomly switched gene fitness effect scores between Chronos and CERES, calculated the difference between their metrics, and repeated 1,000 times. We then calculated the rank of the absolute value of the observed difference among the absolute values of the permutation differences and used this to determine the empirical *p*-value as described in North *et al(North et al., 2002)*.

Code and Data Availability

The code for Chronos is available as a Python package at https://github.com/broadinstitute/chronos. Supporting data and the code used to generate the panels and analyses in this manuscript are available on Figshare at https://doi.org/10.6084/m9.figshare.14067047.

Acknowledgements

Competing Interests

F.V. receives research support from Novo Ventures. D.E.R. receives research funding from members of the Functional Genomics Consortium (Abbvie, Bristol-Myers Squibb, Jannsen, Merck, Vir), and is a director of Addgene, Inc. A.T. is a consultant for Tango Therapeutics, Cedilla Therapeutics, and Foghorn Therapeutics. W.C.H. is a consultant for ThermoFisher, Solasta, MPM Capital, iTeos, Frontier Medicines, Tyra Biosciences, RAPPTA Therapeutics, KSQ Therapeutics, Jubilant Therapeutics and Paraxel.

Aguirre, A. J., Meyers, R. M., Weir, B. A., Vazquez, F., Zhang, C. Z., Ben-David, U., Cook, A., Ha, G., Harrington, W. F., Doshi, M. B., & Others. (n.d.). *Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discov. 2016; 6:*

914--929. doi: 10.1158/2159-8290. CD-16-0154.[PMC free article][PubMed][CrossRef][Google Scholar].

Allen, F., Behan, F., Khodak, A., Iorio, F., Yusa, K., Garnett, M., & Parts, L. (2019). JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome Research*, *29*(3), 464–471.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106.

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., van der Meer, D., Dow, D., Buser-Doepner, C., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516.

Boyle, E. A., Pritchard, J. K., & Greenleaf, W. J. (2018). High-resolution mapping of cancer cell networks using co-functional interactions. *Molecular Systems Biology*, *14*(12). https://doi.org/10.15252/msb.20188594

Chakravarty, D., Gao, J., Phillips, S. M., Kundra, R., Zhang, H., Wang, J., Rudolph, J. E., Yaeger, R., Soumerai, T., Nissan, M. H., Chang, M. T., Chandarlapaty, S., Traina, T. A., Paik, P. K., Ho, A. L., Hantash, F. M., Grupe, A., Baxi, S. S., Callahan, M. K., … Schultz, N. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology*, *2017*. https://doi.org/10.1200/PO.17.00011

Dempster, J. M., Rossen, J., Kazachkova, M., Pan, J., Kugener, G., Root, D. E., & Tsherniak, A. (2019). Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. In *bioRxiv* (p. 720243). https://doi.org/10.1101/720243

DepMap, B. (2019). *Project SCORE processed with CERES* [Data set]. https://doi.org/10.6084/m9.figshare.9116732.v1

DepMap, B. (2020a). *DepMap 20Q2 Public* [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.12280541.V4

DepMap, B. (2020b). *DepMap 20Q4 Public* [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.13237076.V2

De Weck, A., Golji, J., Jones, M. D., Korn, J. M., Billy, E., McDonald, E. R., III, Schmelzle, T., Bitter, H., & Kauffmann, A. (2018). Correction of copy number induced false positives in CRISPR screens. *PLoS Computational Biology*, *14*(7), e1006279.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191.

Fortin, J.-P., Tan, J., Gascoigne, K. E., Haverty, P. M., Forrest, W. F., Costa, M. R., & Martin, S. E. (2019). Multiple-gene targeting and mismatch tolerance can confound analysis of genome-wide pooled CRISPR screens. *Genome Biology*, *20*(1), 21.

Goncalves, E., Behan, F. M., Louzada, S., & Arnol, D. (2018). Tandem duplications lead to loss of fitness effects in CRISPR-Cas9 data. *BioRxiv*. https://www.biorxiv.org/content/10.1101/325076v1.abstract

Hart, T., & Moffat, J. (2016). BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics*, *17*, 164.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832.

Iorio, F., Behan, F. M., Gonçalves, E., Bhosle, S. G., Chen, E., Shepherd, R., Beaver, C., Ansari, R., Pooley, R., Wilkinson, P., Harper, S., Butler, A. P., Stronach, E. A.,

Saez-Rodriguez, J., Yusa, K., & Garnett, M. J. (2018). Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. *BMC Genomics*, *19*(1), 604.

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* , *8*(1), 118–127.

König, R., Chiang, C.-Y., Tu, B. P., Yan, S. F., DeJesus, P. D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., & Chanda, S. K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. *Nature Methods*, *4*(10), 847–849.

Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J. S., Brown, M., & Liu, X. S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology*, *16*, 281.

Luo, B., Cheung, H. W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J. S., Beroukhim, R., Weir, B. A., Mermel, C., Barbie, D. A., Awad, T., Zhou, X., Nguyen, T., Piqani, B., Li, C., Golub, T. R., Meyerson, M., … Root, D. E. (2008). Highly parallel identification of essential genes in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(51), 20380–20385.

Marcotte, R., Sayad, A., Brown, K. R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J. E., Bader, G. D., Mills, G. B., Pe'er, D., Moffat, J., & Neel, B. G. (2016). Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell*, *164*(1-2), 293–309.

Ma, Y., Creanga, A., Lum, L., & Beachy, P. A. (2006). Prevalence of off-target effects in Drosophila RNA interference screens. *Nature*, *443*(7109), 359–363.

McFarland, J. M., Ho, Z. V., Kugener, G., Dempster, J. M., Montgomery, P. G., Bryan, J. G., Krill-Burger, J. M., Green, T. M., Vazquez, F., Boehm, J. S., Golub, T. R., Hahn, W. C., Root, D. E., & Tsherniak, A. (2018). Improved estimation of cancer dependencies from

large-scale RNAi screens using model-based normalization and data integration. *Nature Communications*, *9*(1), 4610.

Meyers, R. M., Bryan, J. G., McFarland, J. M., Weir, B. A., Sizemore, A. E., Xu, H., Dharia, N. V., Montgomery, P. G., Cowley, G. S., Pantel, S., Goodale, A., Lee, Y., Ali, L. D., Jiang, G., Lubonja, R., Harrington, W. F., Strickland, M., Wu, T., Hawes, D. C., … Tsherniak, A. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature Genetics*, *49*(12), 1779–1784.

Michlits, G., Hubmann, M., Wu, S.-H., Vainorius, G., Budusan, E., Zhuk, S., Burkard, T. R., Novatchkova, M., Aichinger, M., Lu, Y., Reece-Hoyes, J., Nitsch, R., Schramek, D., Hoepfner, D., & Elling, U. (2017). CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nature Methods*, *14*(12), 1191–1197.

Michlits, G., Jude, J., Hinterndorfer, M., de Almeida, M., Vainorius, G., Hubmann, M., Neumann, T., Schleiffer, A., Burkard, T. R., Fellner, M., Gijsbertsen, M., Traunbauer, A., Zuber, J., & Elling, U. (2020). Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nature Methods*, *17*(7), 708–716.

Najm, F. J., Strand, C., Donovan, K. F., Hegde, M., Sanson, K. R., Vaimberg, E. W., Sullender, M. E., Hartenian, E., Kalani, Z., Fusi, N., Listgarten, J., Younger, S. T., Bernstein, B. E., Root, D. E., & Doench, J. G. (2017). Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nature Biotechnology*, *36*(2), 179–189.

North, B. V., Curtis, D., & Sham, P. C. (2002). A note on the calculation of empirical P values from Monte Carlo procedures. *American Journal of Human Genetics*, *71*(2), 439–441.

Pacini, C., Dempster, J. M., Gonçalves, E., Najgebauer, H., Karakoc, E., van der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., McFarland, J. M., Garnett, M. J., Tsherniak, A., & Iorio, F. (2020). Integrated cross-study datasets of genetic dependencies in cancer. In

bioRxiv (p. 2020.05.22.110247). https://doi.org/10.1101/2020.05.22.110247

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140.

Shi, J., Wang, E., Milazzo, J. P., Wang, Z., Kinney, J. B., & Vakoc, C. R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nature Biotechnology*, *33*(6), 661–667.

Tálas, A., Huszár, K., Kulcsár, P. I., Varga, J. K., Varga, É., Tóth, E., Welker, Z., Erdős, G., Pach, P. F., Welker, Á., Györgypál, Z., Tusnády, G. E., & Welker, E. (2021). A method for characterizing Cas9 variants via a one-million target sequence library of self-targeting sgRNAs. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1220

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., Gill, S., Harrington, W. F., Pantel, S., Krill-Burger, J. M., Meyers, R. M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W. F. J., Howell, S., Merkel, E., … Hahn, W. C. (2017). Defining a Cancer Dependency Map. *Cell*, *170*(3), 564–576.e16.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., Gozdecka, M., Ohnishi, S., Cooper, J., Patel, M., McKerrell, T., Chen, B., Domingues, A. F., Gallipoli, P., Teichmann, S., … Yusa, K. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Reports*, *17*(4), 1193–1205.

Wu, A., Xiao, T., Fei, T., Shirley Liu, X., & Li, W. (n.d.). *Reducing False Positives in CRISPR/Cas9 Screens from Copy Number Variations*. https://doi.org/10.1101/247031

Yu, J., Silva, J., & Califano, A. (2016). ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics* , *32*(2),

260–267.