Opinion

Synthesizing genome regulation data with vote-counting

Martin Fischer 1,* and Steve Hoffmann 1,*

The increasing availability of high-throughput datasets allows amalgamating research information across a large body of genome regulation studies. Given the recent success of meta-analyses on transcriptional regulators, epigenetic marks, and enhancer:gene associations, we expect that such surveys will continue to provide novel and reproducible insights. However, meta-analyses are severely hampered by the diversity of available data, concurring protocols, an eclectic amount of bioinformatics tools, and myriads of conceivable parameter combinations. Such factors can easily bar life scientists from synthesizing omics data and substantially curb their interpretability. Despite statistical challenges of the method, we would like to emphasize the advantages of joining data from different sources through vote-counting and showcase examples that achieve a simple but highly intuitive data integration.

The curse of omics-data

Mammalian genomes comprise billions of DNA base pairs and thousands of genes. Fueled by increasingly cheap technologies, research into these complex genomes’ organization, function, and regulation has triggered an avalanche of high-throughput experiments [1,2]. On the flip side, the deluge of data and data types requires increasingly sophisticated integration and in silico analysis techniques. Typical research projects frequently involve combinations of experiments to profile the transcriptome, DNA binding landscapes of transcription factors, epigenetic marks, or maps of enhancer:gene associations. Among those protocols, gene expression experiments assume a unique role. Because of the comparably large body of pre-existing knowledge, transcriptomic differences are often easier to interpret and can be linked better to phenotypic variation (e.g., between species, tissues, and single cells). By contrast, the interpretation of methylation differences, for example, is usually less straightforward.

However, even for transcriptomic data, knowledge extraction may become complicated when large sets of (differentially) expressed genes (DEGs) need to be analyzed. Sources of variability, technical and biological, further add to this problem [3]. Technical variability is typically hard to control and arises from differences of reagents, measurement systems, or the handling of samples. However, biological variability from patient to patient, animal to animal, and cell to cell is also a well-known effect partially reflecting gene expression stochasticity [4]. These phenomena are well documented to impede reproducibility for both microarray [5–7] and sequencing technologies [8,9].

While the experimental design should adequately address these problems [e.g., by taking group-specific sample similarities into account (blocking) or increasing sample sizes], this is often not achieved in practice. The balancing act between financial pressures and good experimental planning (e.g., sufficient replicate numbers, suitable sequencing depth, or meaningful validations) can easily be thrown off balance. Yet, even in studies with optimal designs, the results of high-

Highlights

An avalanche of omics data generated by multiple high-throughput experiments provides an opportunity to validate results and to make new discoveries.

While typically regarded as a measure of last resort for meta-analyses, naive vote-counting approaches can have considerable advantages when synthesizing genome regulation data and their simplicity enables a more significant number of researchers to use meta-analyses.

Powerful genomic signatures can be extracted via meta-analyses. Vote-counting-based procedures enable the user to score and rank the results to intuitively balance the tradeoff between specificity and sensitivity.

Vote-counting-based scores and ranks allow for a swift integration of diverse data types (e.g., ChIP-seq-derived transcription factor binding data and RNA-seq-derived differential gene expression data).

1Computational Biology Group, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstraße 11, 07745 Jena, Germany

*Correspondence: Martin.Fischer@leibniz-fli.de (M. Fischer) and Steve.Hoffmann@leibniz-fli.de (S. Hoffmann).
throughput data analysis can be misleading [10]. There is a widespread consensus that it is necessary to obtain independent verification. The large and increasing number of published high-throughput studies provides an opportunity to validate results and improve power by data integration, for example, using meta-analyses.

**Meta-analysis approaches**

Naturally, the strength of a scientific assertion increases with the number and quality of studies arriving at the same conclusion. According to Glass [11], meta-analysis refers to the analysis of analyses and may be defined as the statistical investigation of a large collection of results from individual studies for the purpose of integrating the findings. Thus, it is distinct from the primary analysis and the secondary analysis that scrutinizes an existing dataset with new methods. Statistical meta-analysis provides the means to aggregate independent studies to allow statements about common parameters of interest. In principle, meta-analyses aim to obtain improved estimates of unknown effect sizes or to discover recurrent patterns by gathering data from multiple sources. Indeed, if carefully designed and conducted, meta-analyses can provide more accurate estimates than each of its constituent studies [12]. Therefore, meta-analyses are frequently part of systematic literature reviews (i.e., studies that amalgamate evidence from original articles to answer a specific and well-defined question).

However, the benefits of meta-analyses are not limited to better estimates or the mere gain of statistical power. In our opinion, their most critical advantage is the possibility to assess the generalizability of findings made in individual studies and to identify previously unknown subgroups. While meta-analyses are well-established tools in medicine and biology (e.g., to integrate clinical [13,14] or genome-wide association studies [15]), they are quickly gaining momentum in areas where new data have just begun to accumulate (e.g., for single-cell RNA-seq [16] or metabolomics [17] analyses). Notably, jointly analyzing data from different sources also works to the level of methods. By integrating the results of various analysis tools obtained with the same dataset, ‘wisdom of crowds’-approaches can tackle complex questions in the study of gene regulation networks [18].

Formally, meta-analyses can be described in five stages (problem formulation, data collection, data evaluation, data analysis, and interpretation) and we would like to refer to [19] for a detailed and intuitive explanation of the details. In this context, we would like to point out that imprecise formulations, imbalances in the data collection, or a publication bias (insignificant results are usually not published) may impede the validity of a meta-analysis. While generally beneficial, the increasing number of meta-analysis studies conducted in the past two decades have also revealed some critical shortcomings, which, for instance, can lead to discordant results between meta-analyses on the same topic or question [20,21]. In the case of genome regulation studies, however, publications often comprise genome-wide data (e.g., expression values for all genes of a genome), while focusing only on a smaller subset of the results. Moreover, data and resource publications further ameliorate the publication bias problem.

In the optimal case, meta-analyses use effect sizes (i.e., quantitative measures capturing relevant information from each survey to enable integration). Measured in response to some treatment or obtained by comparing conditions, effect sizes are typically given in terms of standardized mean differences, correlations, or fold changes. In addition to being comparable across the original studies, effect sizes should be consistently interpretable across all measures and variables. They shall express the direction and magnitude of the response [19]. Although controversially discussed, we believe that seemingly naive vote-counting approaches [22] may play a more critical role in the future. Vote-counting captures only the direction of an effect [significantly positive (+1),
the timing of their maximal expression [23]. Meta-analyses of multiple transcriptomic datasets studies reporting the significant differential regulation of a given gene versus the number of studies that do not report such an effect for a given experimental treatment. Subsequently, a decision is made using a minimum vote cut-off: if a certain percentage of studies support the differential expression, the gene is deemed to be influenced by the treatment. In the literature, there is a broad consensus that vote-counting procedures are a measure of last resort in cases where the data are incomplete and need to be applied and interpreted with great care (Box 1). Specifically, vote-counting can become necessary when studies either only report the results of hypothesis tests (e.g., P values for DEGs) or the respective conclusions obtained (e.g., a list of significantly dysregulated genes) [19]. Even if the available data are complete and comprise all necessary information to recalculate effect size estimates such as means or correlations, vote-counting can have considerable advantages. Given the sheer size and number of the datasets available today, recalculations for multiple studies can quickly become unfeasible and cumbersome. While vote-counting data need to be interpreted carefully, it provides a most intuitive measure for the number of studies that agree on a specific finding (e.g., the upregulation of a gene following some treatment). Ultimately, the simplicity of vote-counting enables a more significant number of researchers to use meta-analyses.

Synthesizing diverse genome regulation data

With transcriptional regulation by the tumor suppressor p53 serving as a prominent example (Box 2), many meta-analysis studies integrate transcriptomic data obtained from microarray and RNA-seq experiments. Another example of a frequently studied system for which there are many public data available is the oscillating expression during the cell division cycle. The first transcriptome analysis of the cell cycle proposed stratifying DEGs into five groups based on the timing of their maximal expression [23]. Meta-analyses of multiple transcriptomic datasets

Box 1. Limitations of vote-counting meta-analyses

Vote-counting procedures are critically discussed in the literature and there is an accord that they should be applied with caution. Vote-counting yields no information on the magnitude of an effect and there is no way to assess the required homogeneity of effects across the included studies. Because of this lack of knowledge, naive vote-counting studies are frequently regarded as inferior and have sparked controversial discussions [58]. For instance, Friedman points out that the incorporation of low-powered studies can easily lead to the wrong conclusions (i.e., the failure to find a statistically significant result) [58]. This important caveat is owed to the fact that all studies, irrespective of their sample size and statistical precision, have the same influence on the procedure: exactly one vote. In fact, if the average power of the studies used for vote-counting is smaller than the chosen minimum vote cut-off, then the probability that the vote count leads to the right decision approaches zero as more studies are included [59]. Therefore, vote-counting is especially prone to miss small effects if the average study is underpowered. Also, the selection of the cut-off is left to the analyst and may be difficult to justify. Ultimately, the assessment of the biological relevance needs to be done with alternative methods. While effect size procedures have clear advantages and should be given preference, vote-counting may still be an important and legitimate analysis tool [60]. Most obviously, this is the case when effect sizes are not reported and the available raw data do permit the calculation of effect size estimates. In any case, a strong vote-counting result can help researchers to substantiate their research hypothesis or to detect novel interactions.

In genome regulation studies, the situation can be more cumbersome. For example, in semi-quantitative ChIP-seq studies, effect sizes calculated from the number of aligned reads to a genomic locus (e.g., fold enrichments) are hard to compare across laboratory environments (e.g., because of different instruments or protocol variations) and challenging to interpret. Therefore, such studies are designed to merely establish whether a certain signal is present or absent at a specific locus. In our view, performing optimal effect size-based meta-analyses for next-generation sequencing-based experiments would require the rigorous use of spike-ins (usually not available) and a complete recalculation (e.g., read preprocessing, alignment, peak or differential expression calling) based on the raw data if it is available. Such practical considerations underscore the potential usability of badly reputed vote-counting strategies. While it is clear that vote-counting results need to be interpreted carefully, we believe that they are indispensable tools to summarize available data, develop new hypotheses, and trigger validation experiments. Recent work has proposed more sophisticated forms of vote-counting by combining the counts with effect size estimates [90].
Box 2. Example case: transcription factor p53

In recent years, research into the function of the transcription factor and tumor suppressor p53 has led to meta-analyses that have helped address long-standing questions, uncover its regulatory network, and identify nonreproducible results [61–63]. Similar to other transcription factors, numerous genes are up- or downregulated when the activity of p53 is altered. Typically, only a fraction of genes are regulated directly by the transcription factor, while the remainder are controlled indirectly through other means. While transcription factors may facilitate the activation or repression of transcription, the actual effect on a specific gene may depend on the (epi)genomic context, such as cell state or cell type-specific cofactor binding. For many years, p53 was thought to be an ambiguous transcription factor that up- and downregulates genes depending on such a context. However, deciphering this ‘context code’ remained challenging [61]. High-throughput analyses integrating p53 effects on the transcriptome with genome-wide DNA binding data provided contradictory results. While some studies reported that p53 was binding at genes both up- and downregulated by p53 [64,66], other studies indicated that p53 binding did not occur at genes downregulated after p53 induction [66,67]. Ultimately, it required meta-analyses of microarray and RNA-seq data on p53-dependent gene expression in conjunction with p53-ChIP-seq data to clear the mist [24,32,33,35]. The meta-analyses revealed that many p53 binding signals (ChIP-seq) near genes downregulated by p53 were nonreproducible and more likely represented false positives in the respective studies. As a consequence, all meta-analysis studies amounted to similar conclusions: binding of p53 was significantly enriched only at genes upregulated by p53 and depleted at genes downregulated by p53 (see Figure 1 in main text). Thus, meta-analyses helped to reconcile a vital debate and provided a solid basis for the current model that describes p53 as an activator of transcription [35]. Consequently, p53-induced downregulation is an indirect effect (reviewed in [61–63]). A survey of direct p53 targets revealed that many p53 binding events identified in studies of individual genes (low-throughput) were nonreproducible in high-throughput (ChIP-seq) studies [61]. Notably, p53 binding events identified by the method of ChIP followed by end-point PCR (i.e., DNA quantification by gel documentation after a predetermined number of PCR cycles) were particularly unlikely to be validated by ChIP-seq studies, underscoring that end-point PCR is highly prone compared with state-of-the-art real-time PCR measurements.

In addition to transcriptome data, several more types of omics data can be integrated through meta-analyses. The example of p53 (Box 2) shows that DNA binding data of proteins (i.e., ChIP-seq data) can be integrated with meta-analyses. To integrate multiple high-throughput datasets on DNA binding, vote-counting approaches may operate on the level of ‘peaks’ representing genomic intervals (i.e., DNA sequence regions that exhibit significantly enriched protein binding). In this context, the count can be used to represent the number of datasets providing evidence for the sought DNA–protein interaction for each genomic base pair [32–34]. An alternative approach links each transcription factor binding site to proximal genes and then aggregates the number of datasets with significant binding on the gene level [24,35]. A potentially more powerful method to perform meta-analyses on ChIP-seq data is rank aggregation approaches [36,37]. Meta-analyses of ChIP-seq data are facilitated by tools and databases that provide harmonized resources of ChIP-seq datasets. Examples include the Cistrome Data Browser (CistromeDB) [38], ReMap [39], and UniBind [40].
Transcription factors are not the only chromatin-interacting proteins that are analyzed using ChIP-seq. Most datasets on ChIP-based histone modifications and chromatin states (e.g., promoters, enhancers, or insulators) are published as lists of genomic regions. Again, the integration of such data across different studies can help to get a deeper understanding of genomic regulation. A recent study used publicly available data to identify bivalently marked regions (i.e., genomic segments characterized by the co-occurrence of activating and repressing histone marks) present in most healthy human cells [41]. The study found that these frequently bivalent regions are consistently hypermethylated across numerous cancer types and may play a role in tumorigenesis.

Another interesting type of genomic data is concerned with regulatory 3D chromatin interactions. However, due to experimental difficulties and high costs, only a relatively small number of studies charting genome-wide enhancer:gene associations are available today. To predict such interactions also for cell types for which no such data are available, recent integrative approaches combined the available evidence and consolidated multiple data types to substantially improve the prediction of enhancer:gene associations and ultimately provided valuable resources [42,43]. Certainly, such studies can be expected to accelerate the discovery of new biological insights. Subsets of genes or genomic regions exhibiting solid interaction evidence across datasets provide new starting points for the discovery of biomarkers, signaling pathways, or cell states.

Identification of signatures by merging publicly available studies

The interest in reliable biomarkers substantially fueled biomedical research in the past decades. Today, gene expression signatures, for example, are frequently used in basic research and the clinic, as they can help to measure the activity of signaling pathways and to distinguish cell states. Such approaches have been especially fruitful in oncology and important findings were made by leveraging meta-analyses of various cancer types.

Early microarray-based studies, for instance, established that proliferation and cell cycle genes display an increased expression in most cancers and their expression level often correlates with higher tumor grades and worse prognosis [44–46]. A prominent example is the discovery of signatures revolving around the FOXM1 transcription factor. FOXM1 drives the transcription of cell cycle genes that are most strongly expressed during mitosis (including FOXM1 itself) [26,47–49]. In a more recent meta-analysis using microarray and RNA-seq data from ~18,000 samples representing many cancer types, it has been established that FOXM1 and the network it controls correlate strongly with cancer progression across cancer types [50]. Importantly, this example highlights that the improved power provided through meta-analyses can lead to better estimates, as an earlier analysis of 953 samples from four cancer types identified only limited improvement of predictions using molecular data [51]. Meanwhile, the increasing availability of data for specific cancer types allows ever more detailed insights (e.g., on specific signatures of squamous cell carcinomas derived from a comparison of cancers from squamous cell origin with samples from many other cancer types) [52].

Beyond the success of meta-analyses in cancer research, gene signatures are exploited to determine the differentiation status of cells and to measure the activity of signaling pathways, such as MYC or mTORC1. In the past years, many databases and online services collecting and providing easy access to such signatures became available. Among them are the molecular signatures database (MSigDB) [53], Enrichr [54], or the CistromeDB [38].

Meta-analysis-derived scores and ranks

The careful selection of tools, parameters, and thresholds is an essential step for any data analysis. While these factors are critical for the sensitivity, specificity, precision, or accuracy of a
study, they are not always available when public data are used. Even in cases where all relevant information is known, it is often impossible or highly cumbersome to assess the impact of thresholds (and all combinations thereof) chosen by the authors of the original studies. Thus, the overwhelming majority of users are forced to go forward with the authors’ choices. To partly address this problem without dramatically reducing the usability at the same time, the GeneHancer database, for example, provides two complementary collections. The first collection holds all significant enhancer:gene associations detected in one of the summarized experiments, while the second collection is filtered using a vote-counting threshold. The so-called ‘double-elite’ collection is expected to be more specific (but less sensitive), as all associations need to be detected in at least two of the constituent datasets [42]. To this point, the user can decide whether to use the more sensitive collection and accept more false positives or to resort to the more conservative collection and accept more false negatives.

Unfortunately, users are typically left without such choices. We believe that vote-counting can be a powerful tool enabling intuitive analysis strategies if the average statistical power of the underlying studies is sufficient to detect the sought effects. For instance, counts may be used to score and subsequently rank the results of meta-analyses, providing the user with an intuitive measure to balance the tradeoff between specificity and sensitivity. Recently, we presented a meta-analysis that uses a scoring system for p53-dependent gene expression (the so-called ‘p53 Expression Score’) based on vote-counting [24,35,55]. The score indicates whether a gene is up- or downregulated in response to p53 by simply subtracting the counts of datasets supporting one or the other conclusion. In this way, the user is able to rank and select genes based on their response to perform enrichment analyses or motif searches (e.g., by comparing strong vs. weak responders). Crucially, the vote-counting-based expression score may be used to blend in further genomic information. When integrating ChIP-seq experiments, we were able to establish that genes with high expression scores are also frequently bound by p53. However, genes with low or negative scores do not show such associations, leading to a rejection of models that proposed a frequent function of p53 as a direct repressor of gene expression (Figure 1) [24,35].

Figure 1. Integration of vote-counting meta-analyses on transcription factor (TF) binding data and differential gene expression data. Vote-counts of TF (here p53) ChIP-seq datasets (y-axis) and vote-counts of differential gene expression upon TF (here p53) activation (p53 Expression Score; x-axis). In this example (p53), strong evidence for TF binding correlates only with strong evidence of target gene upregulation. Illustration based on data from [24].
In addition to deciphering the potential contribution of single transcription factors to a gene’s expression, rank-based comparisons of the effects of multiple transcription factors helped us to dissect the transcriptional programs that are usually difficult to discern (e.g., the gene regulatory networks of human p53, its mouse ortholog [34], and its human paralog p63 [56]). Albeit very simple, vote-counting approaches can improve the utility of meta-analyses, provide more precise insights into differences in genome regulatory programs, and enable users to get signature gene sets using individual thresholds.

Concluding remarks and future perspectives
The widespread use of high-throughput experiments has led to an unprecedented accumulation of genome, transcriptome, and epigenome data. Current methods and strategies to annotate, store, and meaningfully analyze these data are far from perfect. The publication and organization of biological data according to the FAIR (Findable, Accessible, Interoperable, Reusable) principles [57], such that it is accessible for meta-analyses, requires intensive and permanent efforts. Still, the implementation of the FAIR principles alone will not be sufficient to substantially facilitate the integration of versatile genome regulation data. We believe that it is increasingly critical to provide and maintain more data resources that synthesize available data in an intelligible manner. Simple integration methods can help to improve the interpretability and ultimately lower the bar to make the most of these precious resources.

Genome regulation studies using meta-analysis approaches have provided new signatures for signaling pathways and cell states. They have helped to disentangle global and cell type-specific gene regulatory networks, as well as direct and indirect subnetworks. Moreover, meta-analyses are helping to dissect networks of ortholog and paralog transcription factors. In the future, the development and application of statistical meta-analysis will be of great advantage for the field of genome regulation, which is facing an unprecedented avalanche of data (see Outstanding questions). The mounting demand for easy-to-use methods and resources is illustrated by the fact that it is often mandatory for authors reporting primary analyses to include publicly available data (e.g., for validation purposes), essentially blurring the line between meta-analysis-centered systematic literature review and original research articles.

While vote-counting approaches have substantial limitations (e.g., being particularly prone to not detecting an effect when the power of their underlying studies is too low), we believe that they are particularly useful for quickly sifting through large bodies of data, to identify potential relationships, and to generate or substantiate new genome regulation hypotheses.

Acknowledgments
Funding for open access charge was provided by Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Jena, Germany. The FLI is a member of the Leibniz Association and is financially supported by the Federal Government of Germany and the State of Thuringia.

Declaration of interests
The authors declare no conflict of interest.

References
1. Reuter, J.A. et al. (2015) High-throughput sequencing technologies. Mol Cell 58, 586–597
2. Fan, J.-B. et al. (2006) Highly parallel genomic assays. Nat. Rev. Genet. 7, 632–644
3. Hansen, K.D. et al. (2011) Sequencing technology does not eliminate biological variability. Nat. Biotechnol. 29, 572–573
4. Chalancon, G. et al. (2012) Interplay between gene expression noise and regulatory network architecture. Trends Genet. 28, 221–232
5. Draghici, S. et al. (2006) Reliability and reproducibility issues in DNA microarray measurements. Trends Genet. 22, 101–109

Outstanding questions
How can we better standardize, organize, and annotate genome regulation data to make them findable, accessible, interoperable, and reusable (FAIR) for meta-analyses and validations?

What easy-to-use tools and user interfaces are required to facilitate the synthesis of genome regulation data for life scientists without extensive knowledge of bioinformatics?

Are currently existing methods sufficient to allow a meaningful aggregation of meta-analyses for different data types (e.g., gene expression and transcription factor binding data for the same samples)?

Should we develop more rigorous standards for next-generation sequencing protocols to facilitate cross-study comparisons (e.g., by using spike-ins for gene expression experiments)?
6. Shi, L. et al. (2008) The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. BMC Bioinformatics 9, 510

7. Leek, J.T. et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. 11, 735–743

8. Tarazona, S. et al. (2011) Differential expression in RNA-seq: a matter of depth. Genome Res. 21, 2213–2223

9. Love, M.I. et al. (2016) Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat. Biotechnol. 34, 1299–1301

10. Nuzzo, R. (2014) Scientific method: statistical errors. Nature 505, 150–152

11. Glass, G.V. (1976) Primary, secondary, and meta-analysis of research. Educ. Res. 5, 3–8

12. Sweeney, T.E. et al. (2017) Methods to increase reproducibility in differential gene expression via meta-analysis. Nucleic Acids Res. 45, e1

13. Sack, H.S. et al. (1997) Meta-analyses of randomized controlled trials. N. Engl. J. Med. 336, 450–455

14. Moher, D. et al. (1999) Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Lancet 354, 1800–1801

15. Paragiotou, O.A. et al. (2013) The power of meta-analysis in genome-wide association studies. Annu. Rev. Genomics Hum. Genet. 14, 441–465

16. Butler, A. et al. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420

17. Zhang, J. et al. (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabonomics analysis. Nucleic Acids Res. 46, W489–W494

18. Marbach, D. et al. (2012) Wisdom of crowds for robust gene network inference. Nat. Methods 9, 786

19. Hartung, J. et al. (2008) Statistical Meta-Analysis with Applications. Wiley

20. Jadad, A.R. et al. (1997) A guide to interpreting discordant systematic reviews. Can. Med. Assoc. J. 156, 1411–1416

21. Leppänen, J. et al. (1997) Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N. Engl. J. Med. 337, 536–542

22. Tiao, G.C. et al. (1990) Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res. 40, 3785–3799

23. Whittaker, M.L. et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell 13, 1997–2002

24. Fischer, M. et al. (2018) Integration of TPS3, DREAM, MBM/FOX1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. Nucleic Acids Res. 44, 6070–6088

25. Gádor, É. et al. (2017) Meta-analysis reveals conserved cell cycle transcriptional network across multiple human cell types. BMC Genomics 18, 30

26. Fischer, M. and Müller, G.A. (2017) Cell cycle transcription control: DREAM/MaxB and RB-E2F complexes. Cell. Res. Biochem. Mol. Biol. 52, 639–662

27. Bravard, D. et al. (2011) The evolution of gene expression levels in mammalian organs. Nature 479, 343–348

28. Barbosa-Morais, N.L. et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. Science 338, 1587–1593

29. Merkin, J. et al. (2012) Evolutionary dynamics of gene isoforms: the solution of mammalian tissues. Science 338, 1593–1599

30. Lin, S. et al. (2014) Comparison of the transcriptional landscapes between human and mouse tissues. Proc. Natl. Acad. Sci. U. S. A. 111, 17224–17229

31. Sudmant, P.H. et al. (2015) Meta-analysis of RNA-seq expression across species, tissues and studies. Genome Biol. 16, 287

32. Verfaillie, A. et al. (2016) Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. Genome Res. 26, 882–895

33. Nguyen, T.-A.T. et al. (2018) Revealing a human p53 universe. Nucleic Acids Res. 46, 8153–8167

34. Fischer, M. (2018) Conservation and divergence of the p53 gene regulatory network between mice and humans. Oncogene 38, 4096–4109

35. Fischer, M. et al. (2014) The transcription factor p53: not a repressor, solely an activator. Cell Cycle 13, 3037–3056

36. Innoshita, H. and Aerts, S. (2018) CHG-seq meta-analysis yields high quality training sets for enhancer classification. bioRxiv Published online August 9, 2018, https://doi.org/10.1101/388904

37. Komiyama, S.K. et al. (2003) Meta-analysis of ChIP-seq datasets through the rank aggregation approach. In 2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB), pp. 180–184

38. Zheng, R. et al. (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. 47, D720–D725

39. Hamrall, F. et al. (2022) ReMap 2022: a database of human, mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res. 50, D216–D225

40. Puig, R.R. et al. (2021) UniBind: maps of high-confidence direct TF-DNA interactions across nine species, BMC Genomics 22, 482

41. Berryhill, S.H. et al. (2016) Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. Sci. Rep. 6, 37393

42. Röhler, S. et al. (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database 2017, baw028

43. Hatt, T.A. et al. (2018) FOCS: a novel method for analyzing enhancer and gene activity patterns using an extensive enhancer-promoter map. Genome Biol. 19, 56

44. Chung, C.H. et al. (2002) Molecular portraits and the family tree of cancer. Nat. Genet. 32, 533–540

45. Rhodes, D.R. et al. (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc. Natl. Acad. Sci. U. S. A. 101, 9309–9314

46. Whitfield, M.L. et al. (2006) Common markers of proliferation. Nat. Rev. Cancer 6, 99–106

47. Lecuyer, J. et al. (2005) FoxM1 is required for execution of the mitotic programme and chromosome stability. Nat. Cell Biol. 7, 126–136

48. Sadasivam, S. et al. (2012) The MuXB complex sequentially recruits B-Myc and FoxM1 to promote mitotic gene expression. Genes Dev. 26, 474–489

49. Chen, X. et al. (2013) The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. Mol. Cell. Biol. 33, 227–236

50. Gentles, A.J. et al. (2015) The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. 21, 939–945

51. Yuan, Y. et al. (2014) Assessing the clinical utility of cancer genomic and proteomic data across tumor types. Nat. Biotechnol. 32, 644–652

52. Campbell, J.D. et al. (2018) Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. Cell Rep. 23, 194–212

53. Libe, R. et al. (2015) The molecular signatures database hallmark gene set collection. Cell Syst. 1, 417–425

54. Kuleshov, M.V. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97

55. Fischer, M. et al. (2022) TargetGeneReg 2.0: a comprehensive web-atlas for p53, p63, and cell-cycle-dependent gene regulation. NAR Cancer 4, zoa009

56. Rege, K. et al. (2023) Dissecting the DNA binding landscape and gene regulatory network of p63 and p53. eLife 9, e62286

57. Wilkinson, M.D. et al. (2016) The FARS guiding principles for scientific data management and stewardship. Sci. Data 3, 160018

58. Friedman, L. (2001) Why vote-count reviews don’t count. Biol. Psychiatry 49, 161–162

59. Hedges, L. and Olkin, I. (1985) Vote-counting methods. In Statistical Methods for Meta-Analysis, pp. 49–52, Academic Press

60. Buschman, B.J. and Wang, M.C. (2009) Vote-counting procedures in meta-analysis. In The Handbook of Research Synthesis
61. Fischer, M. (2017) Census and evaluation of p53 target genes. Oncogene 36, 3943–3956
62. Sullivan, K.D. et al. (2018) Mechanisms of transcriptional regulation by p53. Cell Death Differ. 25, 133–143
63. Sammons, M.A. et al. (2020) Tumor suppressor p53: from engaging DNA to target gene regulation. Nucleic Acids Res. 48, 8848–8869
64. Wei, C.L. et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. Cell 124, 207–219
65. Nikulenkov, F. et al. (2012) Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. Cell Death Differ. 19, 1992–2002
66. Kenzelmann Broz, D. et al. (2013) Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. Genes Dev. 27, 1016–1031
67. Schleith, K. et al. (2013) Characterization of the p53 cistrome—DNA binding cooperativity dissects p53’s tumor suppressor functions. PLoS Genet. 9, e1003726
68. Fischer, M. et al. (2022) Coordinating gene expression during the cell cycle. Trends Biochem. Sci. 47, 1099–1102. https://doi.org/10.1016/j.tibs.2022.06.007