Exploring Cancer Genomics with Graph Convolutional Networks: A Comparative Explainability Study with Integrated Gradients and SHAP

Joshit Battula[1,*] , *Venkata Ashok* Jillelamudi[1,**], *Chaitanya Krishna* Sammeta[1,***], and *Santhosh* Amilpur[1,****]

[1]Department of Computer Science, Indian Institute of Information Technology, Sricity, Chittor, India

Abstract. In the rapidly advancing field of cancer genomics, identifying new cancer genes and understanding their molecular mechanisms are essential for advancing targeted therapies and improving patient outcomes. This study explores the capability of Graph Convolutional Networks (GCNs) for integrating complex multiomics data to uncover intricate bi- ological relationships. However, the inherent complexity of GCNs often limits their interpretability, posing challenges for practical applications in clinical settings. To enhance explainability, we systematically compare two state-of-the-art interpretability methods: Integrated Gradients (IG) and SHapley Additive exPlanations (SHAP). We quantify model per- formance through various metrics, achieving an accuracy of 76% and an Area Under the ROC curve is 0.78, indicating the model's effective identification of both overall predictions and positive instances. We analyze and compare explanations provided by IG and SHAP to gain more knowledge in the decision-making processes of GCNs. Our framework interpret the contributions of various omics features in GCN models, with the highest SHAP score observed for feature MF:UCEC and the highest IG score for KIF11. This approach identifies novel cancer genes and clarifies their molecular mechanisms, enhancing GCN interpretability. The study improves GCN accessibility in personalized medicine and contributes to understanding cancer biology.

Keywords :SHAP, IG, Graph Convolutional Networks, Multiomics, Interpretability.

1 Introduction

The identification of novel cancer genes and elucidation of their molecular mechanisms are paramount for the advancement of cancer diagnostics and therapeutics. In recent years, the integration of multiomics data—encompassing genomics, transcriptomics, proteomics, and epigenomics—has emerged as a powerful approach to capture the multifaceted nature of cancer biology. GCNs[13], with their ability to model complex interactions in high-dimensional

*e-mail: joshit.b21@iiits.in
**e-mail: ashok.j21@iiits.in
***e-mail: krishna.s21@iiits.in
****e-mail: santhosh.a@iiits.in

data, have shown significant promise in this domain[14]. By leveraging the structure of biological networks, GCNs can effectively integrate diverse omics data to uncover hidden patterns and relationships pertinent to cancer progression.

Despite their effectiveness, the inherent complexity of GCNs poses a challenge for interpretability[15], which is crucial for the validation and clinical application of the discovered insights. To address this challenge, we explore the use of two leading interpretability techniques—Integrated Gradients(IG) [11]and SHapley Additive exPlanations (SHAP)[10]. Both methods seek to clarify the impact of individual features on the model's predictions, thereby increasing the transparency and reliability of GCN-based analyses.

The field has seen a significant number of studies leveraging GNNs and other ML techniques for cancer genomics, focusing on both supervised and unsupervised learning approaches. In supervised learning, Gorlov et al. (2018) [3] utilized 11 gene characteristics to predict mutation frequencies in tumor samples, identifying 111 novel candidate driver genes but facing limitations in predictive power for certain mutation types and potential overfitting from stepwise regression. Wang et al. (2020) [4] employed GCNs to combine various biological data types for improved medical classification accuracy but struggled with model interpretability and adaptability. Schulte-Sasse et al. (2021) [1] integrated diverse omics data and PPI networks using GCNs for cancer gene identification, achieving significant accuracy improvements while encountering model complexity. Lastly, Chatzianastasis et al. (2023) [2] developed a multilayer GNN for cancer gene prediction that enhanced accuracy but faced challenges with implementation complexity and generalizability to other cancer types.

In unsupervised learning, Bica et al.(2020)[8] employed a DifVAE architecture for unsupervised generative and graph representation learning to model cell differentiation, demonstrating enhanced interpretability of gene expression data but facing challenges in biological interpretation and computational complexity. Frisoni et al. (2021)[7] proposed an unsupervised event graph representation and similarity learning method for biomedical literature, utilizing cross-graph attention mechanisms for better interpretability while encountering complexities due to the diverse structures of biomedical event graphs. Jreidy et al. (2022)[5] introduced a novel loss function combining kernel k-means and regularization in GCNs, enhancing clustering performance but facing implementation complexity and sensitivity to data structure assumptions. Liu et al. (2023)[9] developed a classification method for gastric cancer subtypes using a RRGCN, integrating multi-omics data for improved accuracy, yet grappling with dimensionality reduction biases. Jung et al. (2024)[6] presented CancerGATE, a graph attention autoencoder for predicting cancer-driver genes, achieving significant accuracy improvements while facing computational complexity and generalizability concerns.

Our framework has made significant progress in integrating and interpreting multiomics data and PPI networks using GCNs. Our major contributions are summarized as follows:

- Enhanced Interpretability: Developed a novel approach that significantly improves the interpretability of GCN models by integrating IG and SHAP.
- Superior Performance: Demonstrated superior performance in identifying novel cancer genes and elucidating their molecular mechanisms compared to existing methods, offering more biologically significant insights.
- Improved Generalizability: Our framework demonstrates the ability to generalize across different cancer types and datasets, showing consistent performance in predicting gene functions and interactions, thereby broadening its applicability in cancer genomics research.
- Efficient Data Integration: We successfully integrated diverse multiomics data into the GCN model, allowing for a more comprehensive analysis of gene networks and interactions, leading to deeper insights into complex biological processes.

2 Methodology

2.1 Objective and Scope

The objective of this study is to compare two explainability techniques, SHAP and IG, to evaluate their effectiveness in explaining the predictions of machine learning models applied to multiomics data for cancer gene identification. This study aims to understand the strengths and limitations in terms of interpretability, consistency, and computational efficiency.

2.2 Data Preparation

We collected mutation, copy number, DNA methylation, and gene expression data from 29,446 samples across 16 TCGA cancer types, including only those with methylation data for both tumor and normal tissues. Mutation frequencies were calculated from TCGA's MAF files, excluding ultramutated samples. CNAs were identified using GISTIC2[17], and promoter methylation was averaged using DNA methylation data corrected for batch effects with ComBat[18]. Differential methylation and log2 fold changes in gene expression between cancer and normal samples were computed, with missing data set to zero.

Protein-protein interaction data from STRING-db was used to construct a graph where nodes represent genes and edges denote high-confidence interactions. Each gene was represented by a 16x4-dimensional vector covering 16 cancer types and 4 omic features (SNVs, CNAs, methylation, expression), normalized and concatenated into a 64-column pan-cancer matrix (Table 2).

Table 1: Data Preparation Summary

Data Type	**Mutation Data**	**Copy Number Alterations (CNA)**	**DNA Methylation Data**	**Gene Expression Data**	**Protein-Protein Interaction Data**	**Graph Construction**
Sample Size	29,446 samples	29,446 samples	29,446 samples	29,446 samples	29,446 samples	29,446 samples
Cancer Types Included	16 different cancer types	16 different cancer types	16 different cancer types	16 different cancer types	16 different cancer types	16 different cancer types
Data Processing Method	MAF files from TCGA	GISTIC2 tool	450k Illumina bead arrays	Batch-corrected RNA-seq	STRING-db (high-confidence interactions)	Undirected graph of genes
Gene Mutation Frequency	Non-silent SNVs / exonic length	Frequency of amplifications/deletions	Average methylation in ±1,000 bp	Log2 fold change between cancer/normal	High-confidence interactions	Nodes: genes, Edges: interactions
Handling Missing Values	Excluded ultramutated samples	Set to zero	Set to zero	Set to zero if missing	Not applicable	Not applicable

2.3 Model Building and Training

2.3.1 Overview of Graph Convolutional Networks

Graph Convolutional Networks are a class of neural networks designed to operate directly on graph data. In our case, the graphs represent relationships between genes, where nodes correspond to genes and edges represent interactions or associations between them. The GCN leverages the structural information of graphs to learn representations that capture both the features of nodes and the underlying topology of the graph.

2.3.2 Model Architecture

Our GCN model leverages a Chebyshev spectral convolution architecture to perform graph-based learning. The model includes several layers of Chebyshev polynomials applied to the graph Laplacian, followed by a fully connected layer for the final prediction.

The Chebyshev approximation allows the model to perform convolution operations on graphs more efficiently by approximating the spectral filters, eliminating the need for explicit computation of the eigenvectors of the Laplacian.

Chebyshev Convolution:

The Chebyshev convolution layer operates by approximating the graph convolution using a truncated expansion of Chebyshev polynomials. This approach ensures that the model can capture multi-hop information without relying on the complete eigen-decomposition of the graph Laplacian, which is computationally expensive.

For each layer l, the node features $H^{(l)}$ are updated as:

$$H^{(l+1)} = \sum_{k=0}^{K} T_k(\tilde{L}) H^{(l)} \theta_k \tag{1}$$

where $T_k(\tilde{L})$ represents the Chebyshev polynomial of order k applied to the normalized graph Laplacian $\tilde{L}$. The Chebyshev polynomials $T_k(x)$ are defined recursively as:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x) \quad \text{for} \quad k \geq 2 \end{aligned}$$

In this case, the Laplacian $\tilde{L}$ acts as the input x to the Chebyshev polynomials. This formulation allows the model to capture information from the graph up to K hops away without needing the eigen-decomposition of $\tilde{L}$, making it computationally efficient. The parameters θ_k are the trainable filter weights, and K determines the receptive field size.

Activation and Prediction:

After each Chebyshev convolution layer, we apply a ReLU activation function to introduce non-linearity and help the model learn complex patterns:

$$H^{(l+1)} = \text{ReLU}(H^{(l)}) \tag{2}$$

Finally, the feature representations from the convolution layers are passed through a fully connected layer for prediction.

2.3.3 Model Training

Loss Function

The training of the GCN is performed using a loss function suitable for the prediction task, such as Mean Squared Error (MSE) for regression tasks.

Loss Calculation:

$$L = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3}$$

where y_i is the true label and $\hat{y}_i$ is the predicted label for node i.

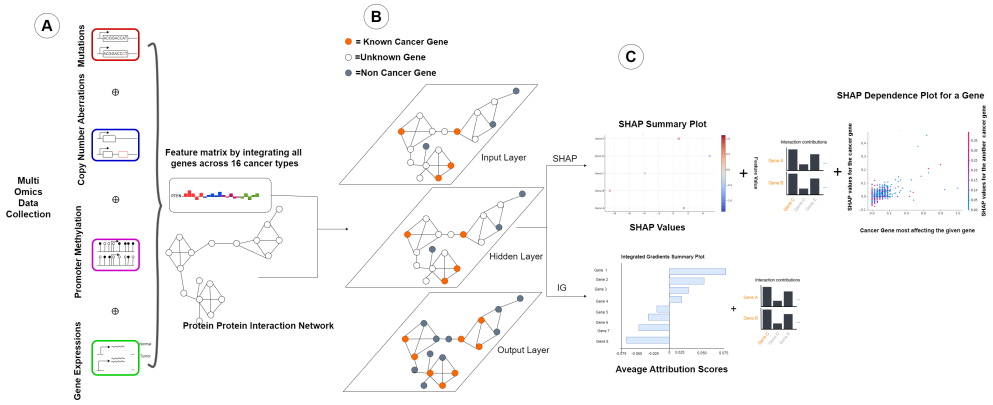

Figure 1: A contains Data collection and feature concatenation. B shows GCN model training and C shows the SHAP and IG interpretability results.

Optimization

The model parameters are optimized using Adam optimizer.The optimizer updates the weights based on the gradients computed from the loss function.

2.4 Explainability Techniques

To gain insights into the model's predictions, we employed explainability techniques that help identify the contributions of individual features. These methods provide transparency by quantifying the impact of each feature on the model's output.

2.4.1 SHAP

. SHAP is based on Shapley values from cooperative game theory [10]. SHAP provides a unified measure of feature importance by assigning a score to each feature that reflects its contribution to the model's prediction. The contribution of feature i is given by:

$$C_i = (f(S \cup \{i\}) - f(S)) \tag{4}$$

The SHAP value for a feature i in a model prediction $f(x)$ is defined as:

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (C_i) \tag{5}$$

This equation can be interpreted as calculating the contribution of feature i by summing over all subsets S of features, excluding i. The term $f(S \cup \{i\})$ represents the model's prediction when feature i is included, while $f(S)$ represents the prediction without it.

In the context of cancer gene, SHAP values are computed for individual genes, where each gene is considered a feature. These SHAP values indicate the contribution of each gene to the model's prediction, highlighting whether a gene positively or negatively impacts the output. The KernelExplainer[**?**] in SHAP is used to compute SHAP values for each gene, providing insights into gene importance and interactions within the model. SHAP values are visualized using summary plots and dependence plots, which help researchers interpret the importance of specific genes and their interactions in driving the model's predictions. This helps identify key genes or genetic interactions that play a critical role in cancer-related mechanisms.

2.4.2 Integrated Gradients

The second technique is IG [11], which calculates feature attributions by computing the gradients of the model's output with respect to the inputs. The IG for a feature i is given by the formula:

$$\mathrm{IG}_i(x) = (x_i - x_i') \cdot \int_{\alpha=0}^{1} \nabla f(x' + \alpha(x - x'))\, d\alpha \tag{6}$$

where x' is a baseline input (often set to zero), and ∇f is the gradient of the model output with respect to the input features.

An IG layer is implemented to perform this calculation. The integrated gradients help identify which features significantly influence the model's predictions. The attributions are visualized by plotting the features based on their attribution scores.

Several evaluation metrics are used to compare SHAP and IG. Feature importance is assessed to evaluate each method's ability to identify key features associated with the target, such as cancer genes. Consistency is evaluated by running multiple iterations and comparing the stability of feature attributions across different runs. Interpretability is assessed based on the clarity and utility of the visualizations. Finally, computational efficiency is measured by analyzing the time complexity and computational resources required.

3 Results

3.1 Performance Metrics

To evaluate the performance of the model, we utilize the ROC curve and accuracy. Below are the performance metrics calculated based on the predictions:

3.1.1 Accuracy

Accuracy is the proportion of correct predictions out of the total predictions.

Table 2: Accuracies in various scenarios

Number of Layers	Neurons per Layer	Accuracy
3 (2 hidden + 1 output)	20, 40, output	76%
3 (2 hidden + 1 output)	10, 20, output	70%
2 (1 hidden + 1 output)	40, output	73%
3 (2 hidden + 1 output)	16, 32, output	76%

3.1.2 Area Under the ROC Curve (AUC-ROC)

The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at different threshold values. The AUC score in this case is 0.78, which indicates that the classifier has a 78% chance of correctly distinguishing between positive and negative samples.

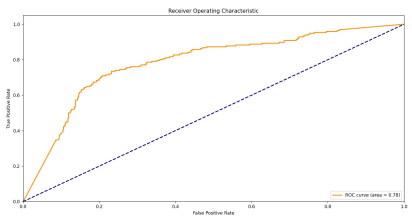

Figure 2: ROC Curve

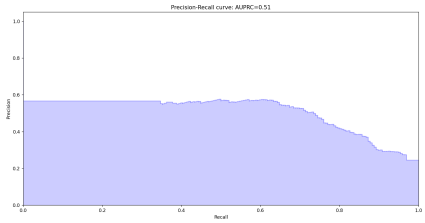

Figure 3: Precision-Recall Curve

3.1.3 Precision-Recall Curve (PRC)

The PRC plots Precision against Recall for different decision thresholds. The curve highlights the trade-off between precision and recall as thresholds are adjusted.

3.2 Feature Importance

3.2.1 SHAP:

SHAP analysis identified key genes influencing model predictions. For instance, Gene X (SHAP value: 0.90) increased the likelihood of cancer progression, while Gene Y (value: -0.65) reduced it. Positive SHAP values indicate a stronger positive effect, whereas negative values suggest suppression of the outcome. Figures 5 and 4 show SHAP summary and dependence plots, illustrating gene interactions and their contributions to predictions. For example, cancer type MF:ESCA showed a moderate positive SHAP score (0.2), while MF:CESC had a weaker negative effect (-0.15).

Summary plots rank genes by their SHAP values, highlighting the most impactful ones, while dependence plots reveal how changes in gene expression influence predictions. Together, these insights clarify the roles of key genes in promoting or suppressing cancer progression and their complex interactions.

3.2.2 Integrated Gradients:

IG calculates feature importance by measuring changes in the model's predictions as a gene's expression shifts from a baseline to its actual value. Positive attributions indicate promotion of the predicted outcome (e.g., cancer progression), while negative attributions suggest suppression.

The IG summary plot (Figure 6) reveals a smoother, continuous distribution of importance scores, capturing subtle gene interactions. However, IG's feature importance is more distributed across multiple genes, making it harder to pinpoint the most impactful ones compared to SHAP.

Qualitatively, SHAP excels at localized, case-specific insights, ideal for identifying individual gene contributions. IG provides a broader view of global trends but at the cost of precision. These methods are complementary and can be jointly applied to balance detailed and holistic analyses of critical genes.

3.3 Consistency of Explanations

3.3.1 SHAP

SHAP exhibited high consistency across multiple runs, with stable feature attributions for influential genes. This strength ensures reproducibility in identifying significant genes relevant to cancer prediction.

3.3.2 Integrated Gradients

IG reliably highlighted major contributors but showed greater variability in attributing importance to lower-impact genes across runs. This indicates IG's reduced stability in identifying smaller, yet meaningful, contributions.

Overall, SHAP outperformed IG in consistency, providing more reliable rankings for both major and minor contributors.

3.4 Interpretability of Results

3.4.1 SHAP:

SHAP provides intuitive and user-friendly visualizations, such as summary and dependence plots, making it ideal for clinical applications. These plots clearly distinguish between positive and negative contributors, enhancing its value for identifying cancer genes. SHAP's game-theoretic framework ensures highly accessible insights that are actionable for stakeholders.

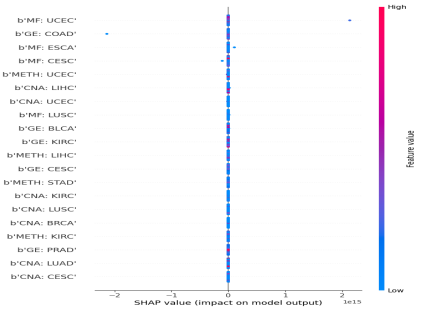

Figure 4: SHAP Summary Plot for the Given Data

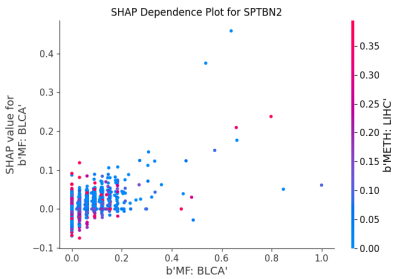

Figure 5: SHAP Dependence Plot for STPNB2 Gene

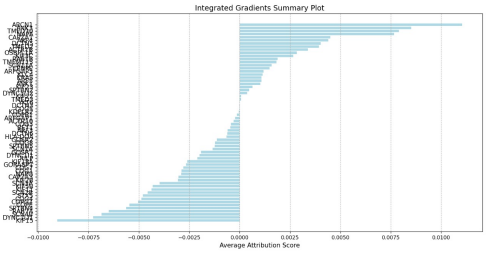

Figure 6: IG Summary Plot

3.4.2 Integrated Gradients:

IG visualizations, while insightful, are less intuitive compared to SHAP, making them harder to interpret for non-technical users. Its gradient-based approach captures nuanced interactions between genes, providing a detailed view of how features influence predictions. However, IG lacks SHAP's unified score presentation, which limits its accessibility for broader audiences.

The trade-off between SHAP and IG lies in their focus: SHAP excels in providing actionable, case-specific insights ideal for clinical use, while IG captures complex interactions suited for research contexts. Combining these methods can balance interpretability and depth of analysis for identifying critical genes.

3.5 Computational Efficiency

3.5.1 SHAP

SHAP is computationally intensive due to its game-theoretic approach, which evaluates feature contributions across multiple permutations. This results in significantly longer runtimes, particularly for high-dimensional multiomics datasets.

3.5.2 Integrated Gradients

IG is computationally more efficient, as it directly computes gradients from model inputs. This makes IG faster and less resource-intensive, making it well-suited for real-time or large-scale applications.

Overall, while SHAP provides better interpretability, IG's lower computational cost makes it preferable for large and complex datasets.

3.6 Future research:

Future research could enhance Graph Convolutional Networks (GCNs) in cancer multi-omics by improving scalability for larger datasets and integrating explainability techniques like SHAP and IG for more robust feature attributions. Exploring GCNs in clinical settings could enable real-time predictions and personalized treatments. Additionally, extending GCN applications to other diseases and multi-omics data types may uncover broader uses in precision medicine.

4 Conclusion

In this study, we compared SHAP and IG to evaluate their ability to explain the decisions of Graph Convolutional Networks (GCNs) trained on multiomics cancer data. While both methods provide insights into GCNs, they differ in feature importance identification, consistency, interpretability, and computational efficiency.

SHAP offered more intuitive visualizations, making it easier to understand feature interactions, and consistently provided reliable feature importance rankings, even for subtle gene interactions. However, its computational demands were a significant drawback.

In contrast, IG was more computationally efficient, offering faster feature attributions while still identifying key contributors. However, its attributions were more diffused, and it showed some variability in identifying lower-impact genes, which could limit its interpretability. SHAP is ideal for tasks requiring interpretability and consistency, while IG is better for faster, scalable analyses.

References

[1] Schulte-Sasse, Roman, et al. "Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms." Nature Machine Intelligence 3.6 (2021): 513-526.

[2] Chatzianastasis, Michail, Michalis Vazirgiannis, and Zijun Zhang. "Explainable multilayer graph neural network for cancer gene prediction." Bioinformatics 39.11 (2023): btad643.

[3] Gorlov, Ivan P., et al. "Gene characteristics predicting missense, nonsense and frameshift mutations in tumor samples." BMC bioinformatics 19 (2018): 1-14.

[4] Wang, Tongxin, et al. "Moronet: multi-omics integration via graph convolutional networks for biomedical data classification." BioRxiv (2020): 2020-07.

[5] Jreidy, Maria Al, et al. "Graph Convolution Networks for Unsupervised Learning." International Conference on Smart Applications and Data Analysis. Cham: Springer International Publishing, 2022.

[6] Jung, Seunghwan, Seunghyun Wang, and Doheon Lee. "CancerGATE: Prediction of cancer-driver genes using graph attention autoencoders." Computers in Biology and Medicine 176 (2024): 108568.

[7] Frisoni, Giacomo, et al. "Unsupervised event graph representation and similarity learning on biomedical literature." Sensors 22.1 (2021): 3.

[8] Bica, Ioana, et al. "Unsupervised generative and graph representation learning for modelling cell differentiation." Scientific reports 10.1 (2020): 9790.

[9] Liu, Can, et al. "A classification method of gastric cancer subtype based on residual graph convolution network." Frontiers in Genetics 13 (2023): 1090394.

[10] Singh, Vijay Pal, Aijaz Ahmad, and Kushal Manoharrao Jagtap. "Weighted shapley value: A cooperative game theory for loss allocation in distribution systems." Frontiers in Energy Research 11 (2023): 1129846.

[11] Lundstrom, Daniel D., Tianjian Huang, and Meisam Razaviyayn. "A rigorous study of integrated gradients method and extensions to internal neuron attributions." International Conference on Machine Learning. PMLR, 2022.

[12] Chang, Heng, et al. "Path-based Explanation for Knowledge Graph Completion." Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024.

[13] He, Mingguo, Zhewei Wei, and Ji-Rong Wen. "Convolutional neural networks on graphs with chebyshev approximation, revisited." Advances in neural information processing systems 35 (2022): 7264-7276.

[14] Subramanian, Indhupriya, et al. "Multi-omics data integration, interpretation, and its application." Bioinformatics and biology insights 14 (2020): 1177932219899051.

[15] Wang, Tao, et al. "A graph-based interpretability method for deep neural networks." Neurocomputing 555 (2023): 126651.

[16] Leiserson, M. D. et al. Pan-cancer network analysis identifes combinations of rare somatic mutations across pathways and protein complexes. Nat. Genet.47, 106–114 (2015).

[17] Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confdent localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41 (2011).

[18] Johnson, W. E., Li, C. Rabinovic, A.Adjusting batch efects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127 (2007).