Data-Adaptive Estimation and Inference in the Analysis of Differential Methylation

for the annual retreat of the *Center for Computational Biology*,
given 18 November 2017

Nima Hejazi

Division of Biostatistics
University of California, Berkeley
stat.berkeley.edu/~nhejazi

nimahejazi.org
twitter/@nshejazi
github/nhejazi

slides: goo.gl/xabp3Q

Preview: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Preview: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Preview: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Preview: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Motivation: Let's meet the data

- Observational study of the impact of disease state on DNA methylation.
- Phenotype-level quantities: 216 subjects, binary disease status (FASD) of each subject, background info on subjects (e.g., sex, age).
- Genomic-level quantities: $\sim 850,000$ CpG sites interrogated using the *Infinium MethylationEPIC BeadChip* by Illumina.
- **Questions**: How do disease status and differential methylation relate? Is a coherent biomarker-type signature detectable?

Motivation: Let's meet the data

- Observational study of the impact of disease state on DNA methylation.
- Phenotype-level quantities: 216 subjects, binary disease status (FASD) of each subject, background info on subjects (e.g., sex, age).
- Genomic-level quantities: $\sim 850,000$ CpG sites interrogated using the *Infinium MethylationEPIC BeadChip* by Illumina.
- **Questions**: How do disease status and differential methylation relate? Is a coherent biomarker-type signature detectable?

Motivation: Let's meet the data

- Observational study of the impact of disease state on DNA methylation.
- Phenotype-level quantities: 216 subjects, binary disease status (FASD) of each subject, background info on subjects (e.g., sex, age).
- Genomic-level quantities: $\sim 850,000$ CpG sites interrogated using the *Infinium MethylationEPIC BeadChip* by Illumina.
- **Questions**: How do disease status and differential methylation relate? Is a coherent biomarker-type signature detectable?

Motivation: Let's meet the data

- Observational study of the impact of disease state on DNA methylation.
- Phenotype-level quantities: 216 subjects, binary disease status (FASD) of each subject, background info on subjects (e.g., sex, age).
- Genomic-level quantities: $\sim 850,000$ CpG sites interrogated using the *Infinium MethylationEPIC BeadChip* by Illumina.
- **Questions**: How do disease status and differential methylation relate? Is a coherent biomarker-type signature detectable?

DNA Methylation

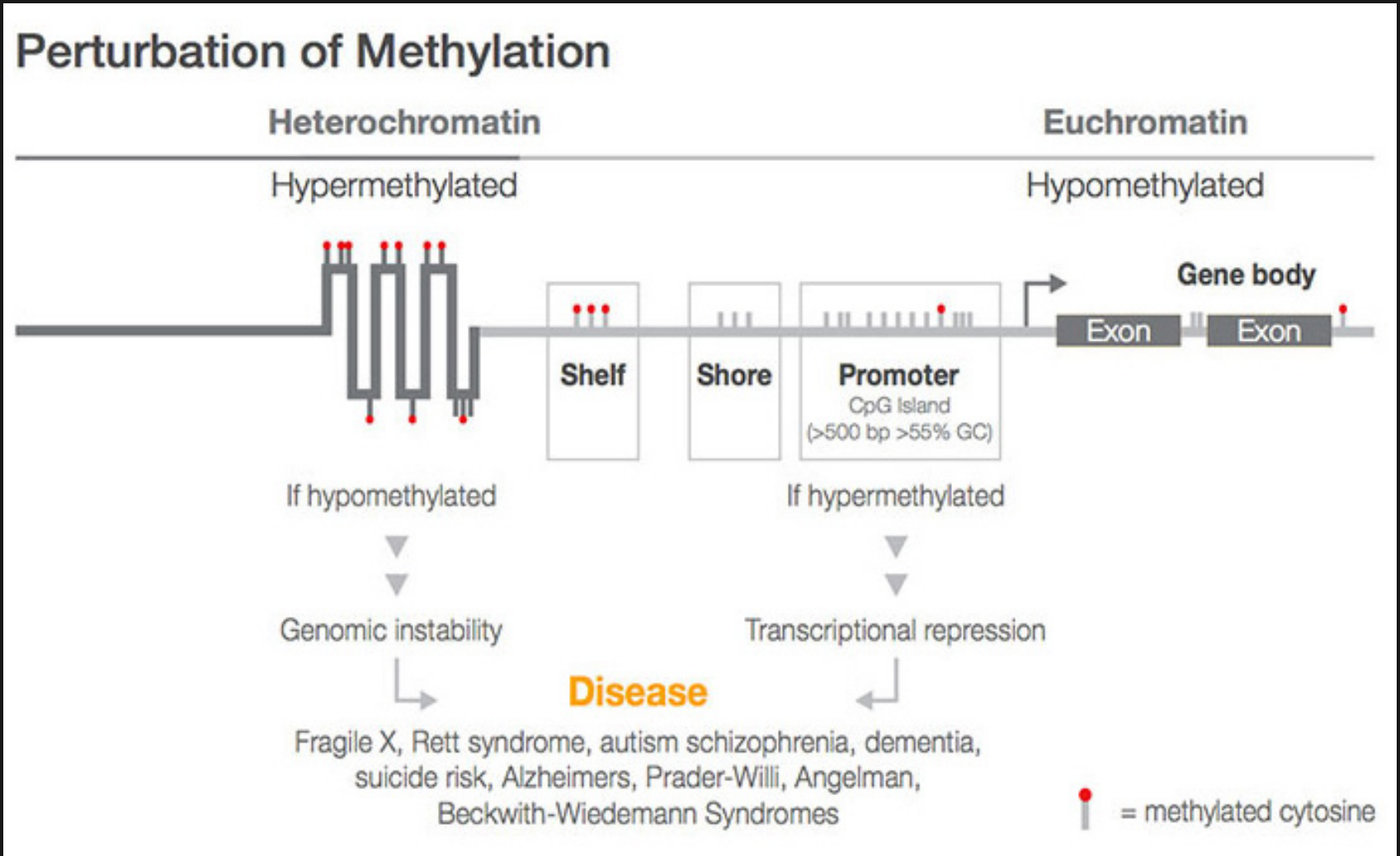

Figure: `https://www.illumina.com/techniques/sequencing/methylation-sequencing.html` (source)

Data analysis? Linear Models!

- Standard operating procedure: For each CpG site ($g = 1, \ldots, G$), fit a linear model:

$$\mathbb{E}[y_g] = X\beta_g$$

- Test the coefficent of interest using a standard t-test:

$$t_g = \frac{\hat{\beta}_g - \beta_{g,H_0}}{s_g}$$

- Such models are a matter of convenience: does $\hat{\beta}_g$ answer our scientific questions? Perhaps not.
- Is consideration being given to whether the data could have been generated by a linear model? Perhaps not.

Data analysis? Linear Models!

- Standard operating procedure: For each CpG site ($g = 1, \ldots, G$), fit a linear model:

$$\mathbb{E}[y_g] = X\beta_g$$

- Test the coefficent of interest using a standard t-test:

$$t_g = \frac{\hat{\beta}_g - \beta_{g,H_0}}{s_g}$$

- Such models are a matter of convenience: does $\hat{\beta}_g$ answer our scientific questions? Perhaps not.
- Is consideration being given to whether the data could have been generated by a linear model? Perhaps not.

Data analysis? Linear Models!

- Standard operating procedure: For each CpG site ($g = 1, \ldots, G$), fit a linear model:

$$\mathbb{E}[y_g] = X\beta_g$$

- Test the coefficent of interest using a standard t-test:

$$t_g = \frac{\hat{\beta}_g - \beta_{g,H_0}}{s_g}$$

- Such models are a matter of convenience: does $\hat{\beta}_g$ answer our scientific questions? Perhaps not.
- Is consideration being given to whether the data could have been generated by a linear model? Perhaps not.

Data analysis? Linear Models!

- Standard operating procedure: For each CpG site ($g = 1, \ldots, G$), fit a linear model:

$$\mathbb{E}[y_g] = X\beta_g$$

- Test the coefficent of interest using a standard t-test:

$$t_g = \frac{\hat{\beta}_g - \beta_{g,H_0}}{s_g}$$

- Such models are a matter of convenience: does $\hat{\beta}_g$ answer our scientific questions? Perhaps not.
- Is consideration being given to whether the data could have been generated by a linear model? Perhaps not.

Motivation: Science Before Statistics

What is the effect of disease status on DNA methylation at a specific CpG site, controlling for the observed methylation status of the neighbors of the given CpG site?

Data analysis? A Data-Adaptive Approach

1. Isolate a subset of CpG sites for which there is cursory evidence of differential methylation.
2. Assign CpG sites into neighborhoods (e.g., bp distance). If there are many neighbors, apply clustering (e.g., *PAM*) to select a subset.
3. Estimate *variable importance measure* (VIM) at each screened CpG site, with disease as intervention (A) and controlling for neighboring CpG sites (W).
4. Apply a variant of the Benjamini & Hochberg method for FDR control, accounting for initial screening.

Data analysis? A Data-Adaptive Approach

1. Isolate a subset of CpG sites for which there is cursory evidence of differential methylation.
2. Assign CpG sites into neighborhoods (e.g., bp distance). If there are many neighbors, apply clustering (e.g., *PAM*) to select a subset.
3. Estimate *variable importance measure* (VIM) at each screened CpG site, with disease as intervention (A) and controlling for neighboring CpG sites (W).
4. Apply a variant of the Benjamini & Hochberg method for FDR control, accounting for initial screening.

Data analysis? A Data-Adaptive Approach

1. Isolate a subset of CpG sites for which there is cursory evidence of differential methylation.
2. Assign CpG sites into neighborhoods (e.g., bp distance). If there are many neighbors, apply clustering (e.g., *PAM*) to select a subset.
3. Estimate *variable importance measure* (VIM) at each screened CpG site, with disease as intervention (A) and controlling for neighboring CpG sites (W).
4. Apply a variant of the Benjamini & Hochberg method for FDR control, accounting for initial screening.

Data analysis? A Data-Adaptive Approach

1. Isolate a subset of CpG sites for which there is cursory evidence of differential methylation.
2. Assign CpG sites into neighborhoods (e.g., bp distance). If there are many neighbors, apply clustering (e.g., *PAM*) to select a subset.
3. Estimate *variable importance measure* (VIM) at each screened CpG site, with disease as intervention (A) and controlling for neighboring CpG sites (W).
4. Apply a variant of the Benjamini & Hochberg method for FDR control, accounting for initial screening.

Pre-Screening — Pick Your Favorite Method

- The estimation procedure is computationally intensive — apply it only to sites that appear promising.
- Consider estimating univariate (linear) regressions of intervention on CpG methylation status. Fast, easy.
- Select CpG sites with a marginal p-value below, say, 0.01. Apply data-adaptive procedure to this subset.
- The modeling assumptions do not matter since the we won't be pursuing inference under such a model.
- Software implementation is extensible. Users are encouraged to add their own. (It's easy!)

Pre-Screening — Pick Your Favorite Method

- The estimation procedure is computationally intensive — apply it only to sites that appear promising.
- Consider estimating univariate (linear) regressions of intervention on CpG methylation status. Fast, easy.
- Select CpG sites with a marginal p-value below, say, 0.01. Apply data-adaptive procedure to this subset.
- The modeling assumptions do not matter since the we won't be pursuing inference under such a model.
- Software implementation is extensible. Users are encouraged to add their own. (It's easy!)

Pre-Screening — Pick Your Favorite Method

- The estimation procedure is computationally intensive — apply it only to sites that appear promising.
- Consider estimating univariate (linear) regressions of intervention on CpG methylation status. Fast, easy.
- Select CpG sites with a marginal p-value below, say, 0.01. Apply data-adaptive procedure to this subset.
- The modeling assumptions do not matter since the we won't be pursuing inference under such a model.
- Software implementation is extensible. Users are encouraged to add their own. (It's easy!)

Pre-Screening — Pick Your Favorite Method

- The estimation procedure is computationally intensive — apply it only to sites that appear promising.
- Consider estimating univariate (linear) regressions of intervention on CpG methylation status. Fast, easy.
- Select CpG sites with a marginal p-value below, say, 0.01. Apply data-adaptive procedure to this subset.
- The modeling assumptions do not matter since the we won't be pursuing inference under such a model.
- Software implementation is extensible. Users are encouraged to add their own. (It's easy!)

Pre-Screening — Pick Your Favorite Method

- The estimation procedure is computationally intensive — apply it only to sites that appear promising.
- Consider estimating univariate (linear) regressions of intervention on CpG methylation status. Fast, easy.
- Select CpG sites with a marginal p-value below, say, 0.01. Apply data-adaptive procedure to this subset.
- The modeling assumptions do not matter since the we won't be pursuing inference under such a model.
- Software implementation is extensible. Users are encouraged to add their own. (It's easy!)

Too Many Neighbors? Clustering

- There are many options: *k*-means, *k*-medoids, etc., as well as many algorithmic solutions.
- For convenience, we use Partitioning Around Medoids (PAM), a well-established algorithm.
- With limited sample sizes, the number of neighboring sites that may be controlled for is limited.
- To faithfully answer the question of interest, choose the neighboring sites that are the most representative.
- This is an *optional* step — it need only be applied when there is a large number of CpG sites in the neighborhood of the target CpG site.

Too Many Neighbors? Clustering

- There are many options: *k*-means, *k*-medoids, etc., as well as many algorithmic solutions.
- For convenience, we use Partitioning Around Medoids (PAM), a well-established algorithm.
- With limited sample sizes, the number of neighboring sites that may be controlled for is limited.
- To faithfully answer the question of interest, choose the neighboring sites that are the most representative.
- This is an *optional* step — it need only be applied when there is a large number of CpG sites in the neighborhood of the target CpG site.

Too Many Neighbors? Clustering

- There are many options: *k*-means, *k*-medoids, etc., as well as many algorithmic solutions.
- For convenience, we use Partitioning Around Medoids (PAM), a well-established algorithm.
- With limited sample sizes, the number of neighboring sites that may be controlled for is limited.
- To faithfully answer the question of interest, choose the neighboring sites that are the most representative.
- This is an *optional* step — it need only be applied when there is a large number of CpG sites in the neighborhood of the target CpG site.

Too Many Neighbors? Clustering

- There are many options: *k*-means, *k*-medoids, etc., as well as many algorithmic solutions.
- For convenience, we use Partitioning Around Medoids (PAM), a well-established algorithm.
- With limited sample sizes, the number of neighboring sites that may be controlled for is limited.
- To faithfully answer the question of interest, choose the neighboring sites that are the most representative.
- This is an *optional* step — it need only be applied when there is a large number of CpG sites in the neighborhood of the target CpG site.

Too Many Neighbors? Clustering

- There are many options: *k*-means, *k*-medoids, etc., as well as many algorithmic solutions.
- For convenience, we use Partitioning Around Medoids (PAM), a well-established algorithm.
- With limited sample sizes, the number of neighboring sites that may be controlled for is limited.
- To faithfully answer the question of interest, choose the neighboring sites that are the most representative.
- This is an *optional* step — it need only be applied when there is a large number of CpG sites in the neighborhood of the target CpG site.

Nonparametric Variable Importance

- Let's consider a simple target parameter: the average treatment effect (ATE):

$$\Psi_g(P_0) = \mathbb{E}_{W,0}[\mathbb{E}_0[Y_g \mid A = 1, W_{-g}] - \mathbb{E}_0[Y_g \mid A = 0, W_{-g}]]$$

- Under certain (untestable) assumptions, interpretable as difference in methylation at site g with intervention and, possibly contrary to fact, the same under no intervention, controlling for neighboring sites.
- Provides a *nonparametric* (model-free) measure for those CpG sites impacted by a discrete intervention.
- Let the choice of parameter be determined by our scientific question of interest.

Nonparametric Variable Importance

- Let's consider a simple target parameter: the average treatment effect (ATE):

$$\Psi_g(P_0) = \mathbb{E}_{W,0}[\mathbb{E}_0[Y_g \mid A = 1, W_{-g}] - \mathbb{E}_0[Y_g \mid A = 0, W_{-g}]]$$

- Under certain (untestable) assumptions, interpretable as difference in methylation at site g with intervention and, possibly contrary to fact, the same under no intervention, controlling for neighboring sites.
- Provides a *nonparametric* (model-free) measure for those CpG sites impacted by a discrete intervention.
- Let the choice of parameter be determined by our scientific question of interest.

Nonparametric Variable Importance

- Let's consider a simple target parameter: the average treatment effect (ATE):

$$\Psi_g(P_0) = \mathbb{E}_{W,0}[\mathbb{E}_0[Y_g \mid A = 1, W_{-g}] - \mathbb{E}_0[Y_g \mid A = 0, W_{-g}]]$$

- Under certain (untestable) assumptions, interpretable as difference in methylation at site g with intervention and, possibly contrary to fact, the same under no intervention, controlling for neighboring sites.
- Provides a *nonparametric* (model-free) measure for those CpG sites impacted by a discrete intervention.
- Let the choice of parameter be determined by our scientific question of interest.

Nonparametric Variable Importance

- Let's consider a simple target parameter: the average treatment effect (ATE):

$$\Psi_g(P_0) = \mathbb{E}_{W,0}[\mathbb{E}_0[Y_g \mid A = 1, W_{-g}] - \mathbb{E}_0[Y_g \mid A = 0, W_{-g}]]$$

- Under certain (untestable) assumptions, interpretable as difference in methylation at site g with intervention and, possibly contrary to fact, the same under no intervention, controlling for neighboring sites.
- Provides a *nonparametric* (model-free) measure for those CpG sites impacted by a discrete intervention.
- Let the choice of parameter be determined by our scientific question of interest.

Target Minimum Loss-Based Estimation

- We use *targeted minimum loss-based estimation* (TMLE), a method for inference in semiparametric infinite-dimensional statistical models.
- No need to specify a functional form or assume that we know the true data-generating distribution.
- **Asymptotic linearity:**

$$\Psi_g(P_n^*) - \Psi_g(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Limiting distribution:**

$$\sqrt{n}(\Psi_n - \Psi) \to N(0, Var(D(P_0)))$$

- **Statistical inference:**

$$\Psi_n \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}}$$

Target Minimum Loss-Based Estimation

- We use *targeted minimum loss-based estimation* (TMLE), a method for inference in semiparametric infinite-dimensional statistical models.
- No need to specify a functional form or assume that we know the true data-generating distribution.
- **Asymptotic linearity:**

$$\Psi_g(P_n^*) - \Psi_g(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Limiting distribution:**

$$\sqrt{n}(\Psi_n - \Psi) \rightarrow N(0, Var(D(P_0)))$$

- **Statistical inference:**

$$\Psi_n \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}}$$

Target Minimum Loss-Based Estimation

- We use *targeted minimum loss-based estimation* (TMLE), a method for inference in semiparametric infinite-dimensional statistical models.
- No need to specify a functional form or assume that we know the true data-generating distribution.
- **Asymptotic linearity:**

$$\Psi_g(P_n^*) - \Psi_g(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Limiting distribution:**

$$\sqrt{n}(\Psi_n - \Psi) \to N(0, Var(D(P_0)))$$

- **Statistical inference:**

$$\Psi_n \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}}$$

Target Minimum Loss-Based Estimation

- We use *targeted minimum loss-based estimation* (TMLE), a method for inference in semiparametric infinite-dimensional statistical models.
- No need to specify a functional form or assume that we know the true data-generating distribution.
- **Asymptotic linearity:**

$$\Psi_g(P_n^*) - \Psi_g(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Limiting distribution:**

$$\sqrt{n}(\Psi_n - \Psi) \to N(0, Var(D(P_0)))$$

- **Statistical inference:**

$$\Psi_n \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}}$$

Target Minimum Loss-Based Estimation

- We use *targeted minimum loss-based estimation* (TMLE), a method for inference in semiparametric infinite-dimensional statistical models.
- No need to specify a functional form or assume that we know the true data-generating distribution.
- **Asymptotic linearity:**

$$\Psi_g(P_n^*) - \Psi_g(P_0) = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Limiting distribution:**

$$\sqrt{n}(\Psi_n - \Psi) \to N(0, Var(D(P_0)))$$

- **Statistical inference:**

$$\Psi_n \pm z_\alpha \cdot \frac{\sigma_n}{\sqrt{n}}$$

Corrections for Multiple Testing

- Multiple testing corrections are critical. Without these, we systematically obtain misleading results.
- The Benjamini & Hochberg procedure for controlling the False Discovery Rate (FDR) is a well-established technique for addressing the multiple testing issue.
- We use a modified BH-FDR procedure to account for the pre-screening step of the proposed algorithm.
- This modified BH-FDR procedure for multi-stage analyses (FDR-MSA) works by adding a p-value of 1.0 for each site that did not pass pre-screening then performs BH-FDR as normal.

Corrections for Multiple Testing

- Multiple testing corrections are critical. Without these, we systematically obtain misleading results.
- The Benjamini & Hochberg procedure for controlling the False Discovery Rate (FDR) is a well-established technique for addressing the multiple testing issue.
- We use a modified BH-FDR procedure to account for the pre-screening step of the proposed algorithm.
- This modified BH-FDR procedure for multi-stage analyses (FDR-MSA) works by adding a p-value of 1.0 for each site that did not pass pre-screening then performs BH-FDR as normal.

Corrections for Multiple Testing

- Multiple testing corrections are critical. Without these, we systematically obtain misleading results.
- The Benjamini & Hochberg procedure for controlling the False Discovery Rate (FDR) is a well-established technique for addressing the multiple testing issue.
- We use a modified BH-FDR procedure to account for the pre-screening step of the proposed algorithm.
- This modified BH-FDR procedure for multi-stage analyses (FDR-MSA) works by adding a p-value of 1.0 for each site that did not pass pre-screening then performs BH-FDR as normal.

Corrections for Multiple Testing

- Multiple testing corrections are critical. Without these, we systematically obtain misleading results.
- The Benjamini & Hochberg procedure for controlling the False Discovery Rate (FDR) is a well-established technique for addressing the multiple testing issue.
- We use a modified BH-FDR procedure to account for the pre-screening step of the proposed algorithm.
- This modified BH-FDR procedure for multi-stage analyses (FDR-MSA) works by adding a p-value of 1.0 for each site that did not pass pre-screening then performs BH-FDR as normal.

Software package: R/methyvim

Figure: `https://bioconductor.org/packages/methyvim`

- Variable importance for discrete interventions.
- Future releases will support continuous interventions.
- Take it for a test drive!

Software package: R/methyvim

Figure: `https://bioconductor.org/packages/methyvim`

- Variable importance for discrete interventions.
- Future releases will support continuous interventions.
- Take it for a test drive!

Software package: R/methyvim

Figure: `https://bioconductor.org/packages/methyvim`

- Variable importance for discrete interventions.
- Future releases will support continuous interventions.
- Take it for a test drive!

Data analysis the methyvim way

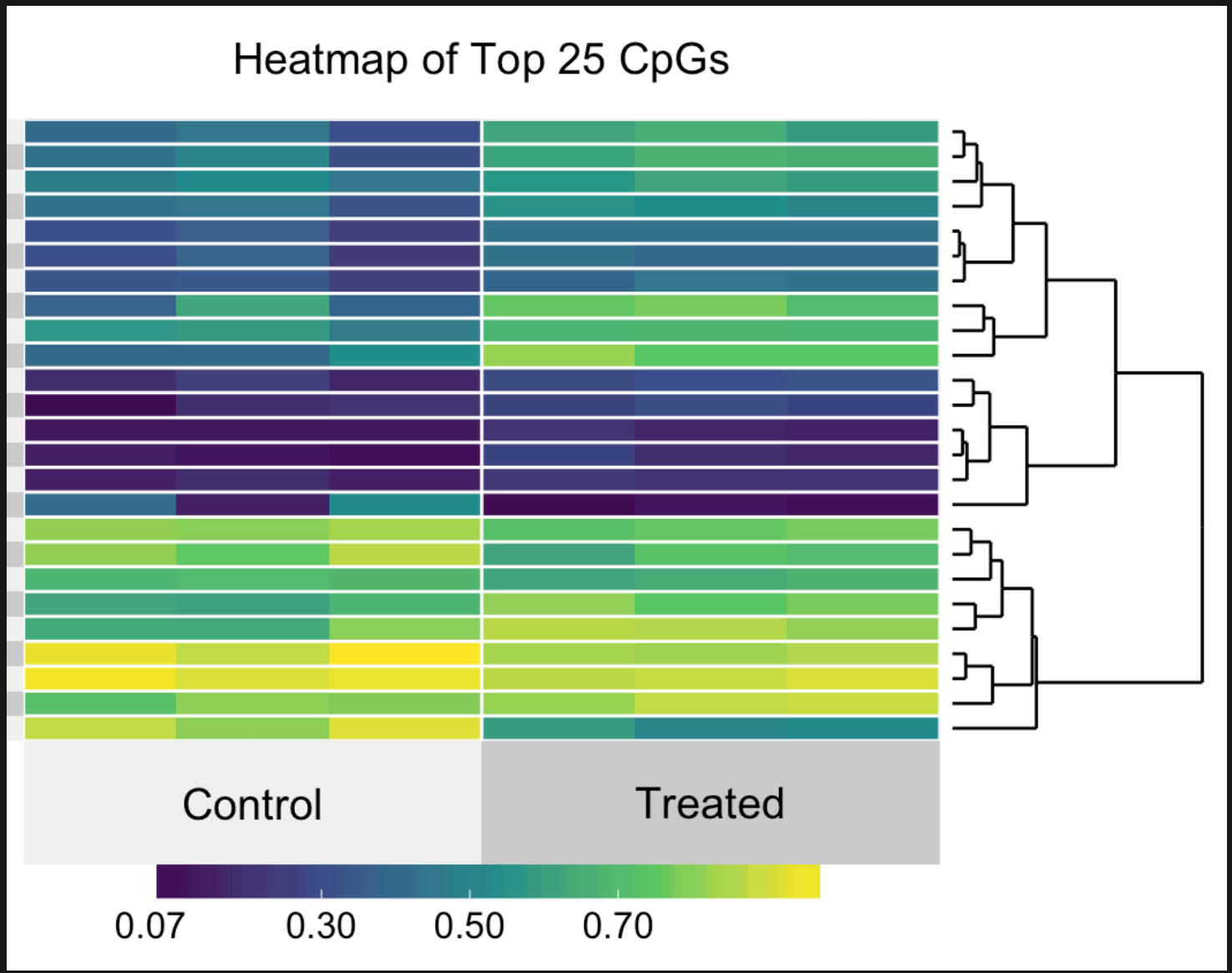

Figure: `http://code.nimahejazi.org/methyvim`

Review: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Review: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Review: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

Review: Summary

- DNA methylation data is *extremely* high-dimensional — we can collect data on 850K genomic sites with modern arrays!
- Normalization and QC are critical components of properly analyzing modern DNA methylation data. There are many choices of technique.
- A relative scarcity of techniques for estimation and inference exists — analyses are often limited to the general linear model.
- Statistical causal inference provides an avenue for answering richer scientific questions, especially when combined with modern advances in machine learning.

References I

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):1–25.

Smyth, G. K. (2005). LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420. Springer Science & Business Media.

References II

Tuglus, C. and van der Laan, M. J. (2009). Modified FDR controlling procedure for multi-stage analyses. *Statistical Applications in Genetics and Molecular Biology*, 8(1):1–15.

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Science & Business Media.

van der Laan, M. J. and Rose, S. (2017). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Science & Business Media.

Acknowledgments

Collaborators:
Mark van der Laan — University of California, Berkeley
Alan Hubbard
Lab of Martyn Smith
Lab of Nina Holland
Rachael Phillips

Funding source:
National Library of Medicine (of NIH): T32LM012417

Thank you.

Slides: goo.gl/JDhSEg

Notes: goo.gl/xabp3Q

Source (repo): goo.gl/m5As73

`stat.berkeley.edu/~nhejazi`

`nimahejazi.org`

`twitter/@nshejazi`

`github/nhejazi`