Supplementary Material for "Extending Classification Algorithms to Case-Control Studies"

June 12, 2019

Within Pair Symmetry of the Conditional Classifiers

As stated in the manuscript, for the 1:1 case-control setting, the conditional-Gaussian naive Bayes classifier and linear discriminant analysis are guaranteed to classify one case and one control per strata.

Gaussian Naive Bayes

If individual 1 in strata i is classified as a case according to the conditional naive Bayes classifier, then individual 2 in strata i will be classified as a control.

Proof. Let x_{ijk} denote feature $k = 1, \ldots, K$ for individual $j = 1, \ldots, g$ in case-control strata $i = 1, \ldots, n$ where g is the group size and n is the number of strata. The response to be modeled is denote y_{ij} where $y_{ij} = 1$ if individual j in strata i is a case and 0 otherwise. For the 1:1 case-control grouping, $g = 2$ and $j \in \{1, 2\}$ and the total number of individuals in the sample is $N = 2n$.

After adjusting for the case control pairing, $x_{i1k} = -x_{i2k}$ and $y_{i1} = 1 - y_{i2}$ for all i and k. It follows that $\hat{\mu}_{1k} = -\hat{\mu}_{2k}$ (see below) and $\sigma^2_{1k} = \sigma^2_{2k}$ for all k (see below). Therefore, $p(x_{i1k}|C_1) = \phi(x_{i1k}; \hat{\mu}_{1k}, \sigma^2_{1k}) = \phi(-x_{i1k}; -\hat{\mu}_{1k}, \sigma^2_{1k}) = \phi(x_{i2k}; \hat{\mu}_{2k}, \sigma^2_{2k}) = p(x_{i2k}|C_2)$. Suppose that individual y_{i1} is predicted to be a case, i.e., $p(\text{Case}) \prod_{k=1}^{K} p(x_{i1k}|\text{Case}) > p(\text{Control}) \prod_{k=1}^{K} p(x_{i1k}|\text{Control})$. Because $p(\text{Case}) = p(\text{Control})$ then

$$\begin{aligned}
&\prod_{k=1}^{K} p(x_{i1k}|\text{Case}) > \prod_{k=1}^{K} p(x_{i1k}|\text{Control}) \\
\implies &\prod_{k=1}^{K} p(x_{i2k}|\text{Control}) > \prod_{k=1}^{K} p(x_{i2k}|\text{Case}) \\
\implies &\hat{y}_{i2} = \text{Control}
\end{aligned}$$

thus the predictions for group i contain one case and one control. This occurs with probability 1 because the "equality" condition has probability 0.

Here we justify the claim that $\hat{\mu}_{1k} = -\hat{\mu}_{2k}$:

$$\begin{aligned}
\hat{\mu}_{1k} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\mathbf{1}(y_{ij}=1)x_{ijk} = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(y_{i1}=1)x_{i1k} + \mathbf{1}(y_{i2}=1)x_{i2k}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(1-y_{i2}=1)(-x_{i2k}) + \mathbf{1}(1-y_{i1}=1)(-x_{i1k})\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(y_{i2}=0)(-x_{i2k}) + \mathbf{1}(y_{i1}=0)(-x_{i1k})\right] = \frac{-1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2}\mathbf{1}(y_{ij}=0)x_{ijk} \\
&= -\hat{\mu}_{2k}.
\end{aligned}$$

Similar techniques can be used to show $\sigma^2_{1k} = \sigma^2_{2k}$.

□

Linear Discriminant Analysis

If individual 1 in strata i is classified as a case according to conditional linear discriminant analysis, then individual 2 in strata i will be classified as a control.

Proof. Using the same notation from the proof for the Gaussian naive Bayes classifier, individual 1 in strata i is classified as a case, i.e., $\hat{y}_{i1} = 1$, if

$$\log\left(\frac{Pr(y_{i1}=1|\boldsymbol{x}_{i1})}{Pr(y_{i1}=0|\boldsymbol{x}_{i1})}\right) = \log\left(\frac{\pi_1\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_1,\boldsymbol{\Sigma})}{\pi_0\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_0,\boldsymbol{\Sigma})}\right) > 0 \tag{1}$$

where π_1 and π_0 are the proportion of individuals in the case and control groups, respectively, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_0$ are the mean vector for the case and control groups, respectively, and $\boldsymbol{\Sigma}$ is the variance covariance matrix that is assumed to be the same in both groups. As before $\boldsymbol{\mu}_0 = -\boldsymbol{\mu}_1$ and the assumption that a single variance covariance matrix $\boldsymbol{\Sigma}$ can be used is guaranteed to be satisfied. Further, due to the balance in cases and controls, $\pi_1 = \pi_0 = 0.5$. Therefore, (**??**) can be rewritten

$$\begin{aligned}
\log\left(\frac{\pi_1\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_1,\boldsymbol{\Sigma})}{\pi_0\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_0,\boldsymbol{\Sigma})}\right) &= \log\left[\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_1,\boldsymbol{\Sigma}) - \log\left(\phi(\boldsymbol{x}_{i1};\boldsymbol{\mu}_0,\boldsymbol{\Sigma}\right]\right. + \log\left(\frac{\pi_1}{\pi_0}\right) \\
&= \boldsymbol{x}_{i1}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) - \frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) + \log(1) \\
&= \boldsymbol{x}_{i1}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)
\end{aligned}$$

because $\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0 = \mathbf{0}$ and $\log(1) = 0$.

Again from (**??**), individual 1 in strata i is classified as a case if $\boldsymbol{x}_{i1}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) > 0$, which implies $-\boldsymbol{x}_{i1}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) = \boldsymbol{x}_{i2}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) < 0$ and therefore individual 2 in strata i is classified as a control.

□

CLR is Special Case of Proposed Methods

We also claimed that the pair corrected and standard CLR methods are mathematically equivalent, which we prove here in two steps. First we show that the conditional likelihood is unaffected by this pair correction.

Proof. Let $\boldsymbol{x}_{ij}$ represent the k-dimensional feature vector for individual $j \in \{1, 2\}$ in stratum $i = 1, \ldots, n$. CLR maximizes the conditional likelihood given by

$$L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) = \prod_{i=1}^{n} P(y_{i1} = 1, y_{i2} = 0|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, y_{i1}+y_{i2} = 1) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})}{\sum_{j=1}^{2} \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{ij})}. \tag{2}$$

For each stratum i, replace the raw feature vectors $\boldsymbol{x}_{ij}$ in (**??**) with the pair corrected feature vectors $\boldsymbol{x}^*_{ij} = \boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot}$ for $j = 1, 2$ gives the exact same likelihood for any pair i:

$$\begin{aligned}
\frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i1})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i2}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i2})} &= \frac{\exp[\boldsymbol{\beta}^\top(\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_{i\cdot})]}{\exp[\boldsymbol{\beta}^\top(\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_{i\cdot})] + \exp[\boldsymbol{\beta}^\top(\boldsymbol{x}_{i1} - \bar{\boldsymbol{x}}_{i\cdot})]} \\
&= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1} - \boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1} - \boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1} - \boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot})} \\
&= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})/\exp(\boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})/\exp(\boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})/\exp(\boldsymbol{\beta}^\top \bar{\boldsymbol{x}}_{i\cdot})} \\
&= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2})}.
\end{aligned}$$

Therefore, CLR is equivalent if the pair corrected or raw data are used to fit the model, i.e., $L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}) = L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*, \boldsymbol{y})$. □

We can now show that standard logistic regression applied to the pair corrected data is equivalent to the standard conditional logistic regression.

Proof. Let $\boldsymbol{x}_{ij}$ represent the k-dimensional feature vector for individual $j \in \{1, 2\}$ in stratum $i = 1, \ldots, n$ and let $\boldsymbol{x}^*_{ij} = \boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{i\cdot}$ represent the pair corrected version of $\boldsymbol{x}_{ij}$ for all i and j. It follows that $\boldsymbol{x}^*_{i1} = -\boldsymbol{x}^*_{i2}$. From above, the contribution of pair i to the likelihood maximized by conditional logistic regression is given by

$$\begin{aligned}
\frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i1}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2})} &= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i1})}{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i1}) + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i2})} \\
&= \frac{1}{1 + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i2})/\exp(\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i1})} \\
&= \frac{1}{1 + \exp[\boldsymbol{\beta}^\top(\boldsymbol{x}^*_{i2} - \boldsymbol{x}^*_{i1})]} \\
&= \frac{1}{1 + \exp(2\boldsymbol{\beta}^\top \boldsymbol{x}^*_{i2})}
\end{aligned}$$

The likelihood of standard logistic regression (without an intercept) for the pair corrected data is given by

$$L_{LR}(\boldsymbol{\beta}|\boldsymbol{X}^*, \boldsymbol{y}) = \prod_{i=1}^{n}\prod_{j=1}^{2}\left(\frac{1}{1+\exp(-\boldsymbol{\beta}^\top \boldsymbol{x}_{ij}^*)}\right)^{y_{ij}}\left(\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{ij}^*)}\right)^{1-y_{ij}}.$$

As defined, $y_{i1} = 1$ and $y_{i2} = 0$ for all i. Therefore, the intercept free logistic regression likelihood can be written

$$\begin{aligned}
L_{LR}(\boldsymbol{\beta}|\boldsymbol{X}^*, \boldsymbol{y}) &= \prod_{i=1}^{n}\left(\frac{1}{1+\exp(-\boldsymbol{\beta}^\top \boldsymbol{x}_{i1}^*)}\right)\left(\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)}\right)\\
&= \prod_{i=1}^{n}\left(\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)}\right)\left(\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)}\right)\\
&= \prod_{i=1}^{n}\left(\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)}\right)^2\\
&= \left(\prod_{i=1}^{n}\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)}\right)^2\\
&= \left[L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*/2, \boldsymbol{y})\right]^2
\end{aligned}$$

Suppose $\widehat{\boldsymbol{\beta}}$ maximizes the conditional logistic likelihood, that is

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}} L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*, \boldsymbol{y}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}}\prod_{i=1}^{n}\frac{1}{1+\exp(2\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)} \text{ then}$$

$$2\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}}\prod_{i=1}^{n}\frac{1}{1+\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}} L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*/2, \boldsymbol{y}).$$

Because $1/[1+\exp(2\boldsymbol{\beta}^\top \boldsymbol{x}_{i2}^*)] \geq 0$ for all i, then the vector $\boldsymbol{\beta}$ that maximizes the likelihood function, also maximizes the square of the likelihood function, i.e.,

$$2\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}}\, L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*/2, \boldsymbol{y}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}}\, L_{CLR}(\boldsymbol{\beta}|\boldsymbol{X}^*/2, \boldsymbol{y})^2 = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\text{argmax}}\, L_{LR}(\boldsymbol{\beta}|\boldsymbol{X}^*, \boldsymbol{y}).$$

Therefore, if $\boldsymbol{\beta}$ is the maximum likelihood estimator (MLE) for conditional logistic regression, then $2\boldsymbol{\beta}$ is the MLE for the logistic regression of the same data after centering each pair and setting the intercept to be 0.

□

From these two proof we can conclude that the results obtained by fitting a standard conditional logistic regression to a dataset can be replicated exactly by fitting a standard logistic regression to the data corrected as we proposed in this manuscript (and scaling the regression coefficients appropriately). As such, we conclude that conditional logistic regression is a special case of the larger class of classification algorithms we proposed in this manuscript.

The TEDDY Study Group

Colorado Clinical Center: Marian Rewers, M.D., Ph.D., PI[1,4,5,6,10,11], Kimberly Bautista[12], Judith Baxter[9,10,12,15], Daniel Felipe-Morales, Kimberly Driscoll, Ph.D.[9], Brigitte I. Frohnert, M.D.[2,14], Marisa Gallant, M.D.[13], Patricia Gesualdo[2,6,12,14,15], Michelle Hoffman[12,13,14], Rachel Karban[12], Edwin Liu, M.D.[13], Jill Norris, Ph.D.[2,3,12], Adela Samper-Imaz, Andrea Steck, M.D.[3,14], Kathleen Waugh[6,7,12,15], Hali Wright[12]. University of Colorado, Anschutz Medical Campus, Barbara Davis Center for Childhood Diabetes.

Finland Clinical Center: Jorma Toppari, M.D., Ph.D., PI[¥^1,4,11,14], Olli G. Simell, M.D., Ph.D., Annika Adamsson, Ph.D.[^12], Suvi Ahonen[*±§], Heikki Hyöty, M.D., Ph.D.[*±6], Jorma Ilonen, M.D., Ph.D.[¥¶3], Sanna Jokipuu[^], Leena Karlsson[^], Miia Kähönen[µ¤], Mikael Knip, M.D., Ph.D.[*±5], Mirva Koreasalo[*±§2], Kalle Kurppa, M.D., Ph.D.[*±13], Tiina Latva-aho[µ¤], Maria Lönnrot, M.D., Ph.D.[*±6], Markus Mattila*, Elina Mäntymäki[^], Katja Multasuo[µ¤], Tiina Niininen[±*12], Sari Niinistö[±§2], Mia Nyblom[*±], Sami Oikarinen, Ph.D.[*±], Paula Ollikainen[µ¤] , Petra Rajala[^], Jenna Rautanen[±§], Anne Riikonen[*±§], Minna Romo[^], Suvi Ruohonen[^], Juulia Rönkä[µ¤], Satu Simell, M.D., Ph.D.[¥13], Tuula Simell, Ph.D.[¥12], Maija Sjöberg[¥^12,14], Aino Stenius[µ¤12], Sini Vainionpää[^], Eeva Varjonen[¥^12], Riitta Veijola, M.D., Ph.D.[µ¤14], Suvi M. Virtanen, M.D., Ph.D.[*±§2], Mari Vähä-Mäkilä[^], Mari Åkerlund[*±§], Katri Lindfors, Ph.D.[*13] [¥]University of Turku, *University of Tampere, [µ]University of Oulu, ^Turku University Hospital, Hospital District of Southwest Finland, [±]Tampere University Hospital, [¤]Oulu University Hospital, §National Institute for Health and Welfare, Finland, [¶]University of Kuopio.

Georgia/Florida Clinical Center: Jin-Xiong She, Ph.D., PI[1,3,4,11], Desmond Schatz, M.D.[*4,5,7,8], Diane Hopkins[12], Leigh Steed[12,13,14,15], Jennifer Bryant, Janey Adams[*12], Katherine Silvis[2], Michael Haller, M.D.[*14], Melissa Gardiner, Richard McIndoe, Ph.D., Ashok Sharma, Stephen W. Anderson, M.D.[^], Laura Jacobsen, M.D.[*14], John Marks, DHSc.*, P.D. Towe*. Center for Biotechnology and Genomic Medicine, Augusta University. *University of Florida, ^Pediatric Endocrine Associates, Atlanta.

Germany Clinical Center: Anette G. Ziegler, M.D., PI[1,3,4,11], Andreas Beyerlein, Ph.D.[2], Ezio Bonifacio Ph.D.[*5], Anita Gavrisan, Cigdem Gezginci, Anja Heublein, Michael Hummel, M.D.[13], Sandra Hummel, Ph.D.[2], Annette Knopff[7], Charlotte Koch, Sibylle Koletzko, M.D.[¶13], Claudia Ramminger, Roswith Roth, Ph.D.[9], Marlon Scholz, Joanna Stock[9,12,14], Katharina Warncke, M.D.[14], Lorena Wendel, Christiane Winkler, Ph.D.[2,12,15]. Forschergruppe Diabetes e.V. and Institute of Diabetes Research, Helmholtz Zentrum München, Forschergruppe Diabetes, and Klinikum rechts der Isar, Technische Universität München. *Center for Regenerative Therapies, TU Dresden, [¶]Dr. von Hauner Children's Hospital, Department of Gastroenterology, Ludwig Maximillians University Munich.

Sweden Clinical Center: Åke Lernmark, Ph.D., PI[1,3,4,5,6,8,10,11,15], Daniel Agardh, M.D., Ph.D.[13], Carin Andrén Aronsson, Ph.D.[2,12,13], Maria Ask, Jenny Bremer, Ulla-Marie Carlsson, Corrado Cilio, Ph.D., M.D.[5], Emelie Ericson-Hallström, Annika Fors, Lina Fransson, Thomas Gard, Rasmus Bennet, Carina Hansson, Susanne Hyberg, Hanna Jisser, Fredrik Johansen, Berglind Jonsdottir, M.D., Ph.D., Silvija Jovic, Helena Elding Larsson, M.D., Ph.D. [6,14], Marielle Lindström, Markus Lundgren, M.D., Ph.D.[14], Maria Månsson-Martinez, Maria Markan, Jessica Melin[12], Zeliha Mestan, Caroline

Nilsson, Karin Ottosson, Kobra Rahmati, Anita Ramelius, Falastin Salami, Sara Sibthorpe, Anette Sjöberg, Birgitta Sjöberg, Carina Törn, Ph.D. [3,15], Anne Wallin, Åsa Wimar[14], Sofie Åberg. Lund University.

Washington Clinical Center: William A. Hagopian, M.D., Ph.D., PI[1,3,4, 5, 6,7,11,13, 14], Michael Killian[6,7,12,13], Claire Cowen Crouch[12,14,15], Jennifer Skidmore[2], Ashley Akramoff, Jana Banjanin, Masumeh Chavoshi, Kayleen Dunson, Rachel Hervey, Rachel Lyons, Arlene Meyer, Denise Mulenga, Jared Radtke, Davey Schmitt, Julie Schwabe, Sarah Zink. Pacific Northwest Research Institute.

Pennsylvania Satellite Center: Dorothy Becker, M.D., Margaret Franciscus, MaryEllen Dalmagro-Elias Smith[2], Ashi Daftary, M.D., Mary Beth Klein, Chrystal Yates. Children's Hospital of Pittsburgh of UPMC.

Data Coordinating Center: Jeffrey P. Krischer, Ph.D.,PI[1,4,5,10,11], Sarah Austin-Gonzalez, Maryouri Avendano, Sandra Baethke, Rasheedah Brown[12,15], Brant Burkhardt, Ph.D.[5,6], Martha Butterworth[2], Joanna Clasen, David Cuthbertson, Christopher Eberhard, Steven Fiske[9], Dena Garcia, Jennifer Garmeson, Veena Gowda, Kathleen Heyman, Belinda Hsiao, Francisco Perez Laras, Hye-Seung Lee, Ph.D.[1,2,13,15], Shu Liu, Xiang Liu, Ph.D.[2,3,9,14], Kristian Lynch, Ph.D. [5,6,9,15], Colleen Maguire, Jamie Malloy, Cristina McCarthy[12,15], Aubrie Merrell, Steven Meulemans, Hemang Parikh, Ph.D.[3], Ryan Quigley, Cassandra Remedios, Chris Shaffer, Laura Smith, Ph.D.[9,12], Susan Smith[12,15], Noah Sulman, Ph.D., Roy Tamura, Ph.D.[1,2,13], Ulla Uusitalo, Ph.D.[2,15], Kendra Vehik, Ph.D.[4,5,6,14,15], Ponni Vijayakandipan, Keith Wood, Jimin Yang, Ph.D., R.D.[2,15]. *Past staff: Michael Abbondondolo, Lori Ballard, David Hadley, Ph.D., Wendy McLeod.* University of South Florida.

Project scientist: Beena Akolkar, Ph.D.[1,3,4,5,6,7,10,11]. National Institutes of Diabetes and Digestive and Kidney Diseases.

Proteomics Laboratory: Richard D. Smith, Ph.D., Thomas O. Metz, Ph.D., Charles Ansong, Ph.D., Bobbie-Jo Webb-Robertson, Ph.D., Hugh D. Mitchell, Ph.D., Ernesto S. Nakayasu, Ph.D., and Wei-Jun Qian, Ph.D. Pacific Northwest National Laboratory.

SNP Laboratory: Stephen S. Rich, Ph.D.[3], Wei-Min Chen, Ph.D.[3], Suna Onengut-Gumuscu, Ph.D.[3], Emily Farber, Rebecca Roche Pickin, Ph.D., Jonathan Davis, Jordan Davis, Dan Gallo, Jessica Bonnie, Paul Campolieto. Center for Public Health Genomics, University of Virginia.

Repository: Sandra Ke, Niveen Mulholland, Ph.D. NIDDK Biosample Repository at Fisher BioServices.

Other contributors: Kasia Bourcier, Ph.D.[5], National Institutes of Allergy and Infectious Diseases. Thomas Briese, Ph.D.[6,15], Columbia University. Suzanne Bennett Johnson, Ph.D.[9,12], Florida State University. Eric Triplett, Ph.D.[6], University of Florida.

Committees:
[1]Ancillary Studies, [2]Diet, [3]Genetics, [4]Human Subjects/Publicity/Publications, [5]Immune Markers, [6]Infectious Agents, [7]Laboratory Implementation, [8]Maternal Studies, [9]Psychosocial, [10]Quality Assurance, [11]Steering, [12]Study Coordinators, [13]Celiac Disease, [14]Clinical Implementation, [15]Quality Assurance Subcommittee on Data Quality.