DISSERTATION

ASSESSMENT OF RAPID EVAPORATIVE IONIZATION MASS SPECTROMETRY

(REIMS) TO CHARACTERIZE BEEF QUALITY AND THE IMPACT OF OVEN

TEMPERATURE AND RELATIVE HUMIDTY ON BEEF

Submitted by

Devin Gredell

Department of Animal Sciences

Doctoral Committee:

    Advisor: Dale Woerner

    Keith Belk
    Terry Engle
    Jessica Prenni
    Adam Heuberger

ABSTRACT

ASSESSMENT OF RAPID EVAPORATIVE IONIZATION MASS SPECTROMETRY (REIMS) TO CHARACTERIZE BEEF QUALITY AND THE IMPACT OF OVEN TEMPERATURE AND RELATIVE HUMIDTY ON BEEF

The objective of experiment 1 was to evaluate the ability of rapid evaporative ionization mass spectrometry (REIMS) to predict beef eating quality characteristics. Striploin sections (5 cm in thickness; N = 292) from 7 beef carcass types (Select, Low Choice, Top Choice, Prime, Dark Cutter, Grass-fed, and Wagyu) were collected to achieve variation in fat content, sensory attributes, tenderness, and production background. Sections were aged for 14 d, fabricated into 2.54 cm thick steaks, and frozen until analysis. Trained descriptive panel rated tenderness, flavor, and juiciness attributes for sensory prediction models. Slice shear force (SSF) and Warner-Bratzler shear force (WBS) values were measured to predict tenderness classifications. A molecular fingerprint of each sample was collected via REIMS to build prediction models. Models were built using 80% of samples that were selected randomly for this purpose and tested for prediction accuracy using the remaining 20%. Partial least squares (PLS) discriminant analysis was used as a dimension reduction technique before building a linear discriminant analysis (LDA) model for classification. When Select and Low Choice samples, as well as Top Choice and Prime samples, were combined, balanced prediction accuracy reached 83.8%. Slice shear force and WBS tenderness classifications (tough vs tender) were predicted with 75.0% and 70.2% accuracy, respectively. Sensory models were built to assign samples into positive and negative classifications based on either all sensory attributes (i.e., tenderness, juiciness, and

ii

flavor) or only flavor attributes. Overall sensory class was predicted with 75.4% accuracy and flavor class with 70.3%. With future fine-tuning, these data suggest that REIMS produces a metabolic fingerprint to provide a method to meaningfully predict numerous beef quality attributes in an on-line application.

The objective of the second study was to evaluate the roles of cooking rate and relative humidity on sensory development of beef strip steaks. Thirty USDA Choice beef strip loins were collected from a commercial packing facility. Each strip loin was cut into steaks and randomly assigned to 1 of 6 cooking methods utilizing 2 oven temperatures (80°C and 204°C) and 3 levels of relative humidity [zero (ZH), mid (MH), and high (HH)]. Cooked steaks were used to evaluate internal and external color, Warner-Bratzler and slice shear force, total collagen content, protein denaturation, and trained sensory ratings. Relative humidity greatly reduced cooking rate, especially at 80°C. Steaks cooked at 80°C-ZH had the greatest ($P < 0.01$) cook loss of all treatments, and cook loss was not affected ($P > 0.05$). Steaks cooked at 80°C-ZH appeared the most ($P < 0.01$) well-done and had the darkest ($P > 0.01$) surface color. Total collagen was greatest ($P < 0.01$) in steaks cooked with ZH, regardless of oven temperature. Myosin denaturation was not affected ($P > 0.05$) by treatment. Increased ($P = 0.02$) sarcoplasmic protein denaturation was observed with ZH and MH, while increased ($P = 0.02$) actin denaturation was observed only with ZH. Oven temperature did not influence ($P > 0.05$) protein denaturation. Trained panelists rated steaks most tender ($P < 0.01$) when cooked at 80°C and with ZH and MH. Humidity did not affect ($P > 0.05$) juiciness at 204°C; however, MH and HH produced a juicier ($P < 0.01$) steak when cooked at 80°C. Humidity hindered ($P < 0.01$) the development of beefy/brothy and brown/grilled flavors but increased ($P = 0.01$) metallic/bloody intensity. Lower

oven temperatures and moderate levels of humidity could be utilized to maximize tenderness, while minimally affecting flavor development.

ACKNOWLEDGEMENTS

As I think about where to begin my acknowledgements, it is incredibly humbling to reflect on all those who have helped me get to this point. There are few things in life a person can truly accomplish on their own and completing a PhD program certainly is not one of them. Throughout my academic career, I have been blessed to have made some of the best friends, learn from some of the best mentors, and study at some of the best animal science programs in the country. It has not been a single experience that has shaped me into the man I am today, but rather a culmination of all the people and opportunities along the way.

I first must thank my family for the love and support they've given me my entire life. As the youngest of four, I've always had an abundance of incredible mentors to look up to. Even though they are still learning what meat science is, that has never stopped them from encouraging me and cheering me through this entire process.

None of my accomplishments at Colorado State would have been possible without the help of my fellow graduate students. I am so grateful that several of you have become some of my most valued friendships. There is no better support group in graduate school than those in it with you. KR, Christy, Josh, Blake, Brenna, Joanna, and Matheus, I cannot thank you all enough for everything and for being the friends that you are. Each of you never hesitated to help when I needed it. To be quite candid, there are days in grad that just suck, but being surrounded by individuals like yourselves made even the longest, hardest, most miserable days enjoyable. Although you may not realize it, each of you have helped me become a better person and a better scientist in one way or another. And thank you to all my fellow students who selflessly helped me throughout the years, through both the good and the bad sensory panel samples.

Thank you, Dr. Woerner, for taking the chance on me and agreeing to advise me the last three years. I truly cannot thank you enough for your mentorship. There are few, if any, people I have met that are as passionate, hard-working, and genuine as yourself. For as quick as you were to acknowledge a job well-done, you were just as quick to let me know where I left room for improvement. I will always be incredibly thankful for the tough love you've given me and never lowering your expectations. At the time, it may not have always been what I wanted to hear, but it was always what I needed to hear. Thank you.

I must also thank the rest of my committee. Dr. Belk, I have learned some of the most critical theoretical lessons about being a scientist from you. You were never exaggerating when you said this program is like boot camp, but it is that mentality that gives us the experiences and opportunities to make us successful in our careers. Dr. Engle, thank you for always being so approachable and willing to help any way you could. Dr. Prenni, approaching all this metabolomics data would have been a much greater task for me if it hadn't been for your guidance. Thank you for always being excited to answer any of my seemingly endless questions. Dr. Heuberger, thank you for your unique perspective. Every time we met, you always had a new way to approach the data that I had overlooked.

Lastly, I need to give one final heartfelt thank you to Dr. Geornaras. I know I was never a "micro" kid. Nevertheless, I have learned things from you that I will career with me throughout my entire career. You taught me the true value of collecting quality data and to never be afraid of starting over if it means saving the integrity of your research. You taught me to always collect data the best way and that the best way isn't always the easiest way. Your integrity is second to none, and I will always value your mentorship and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INRODUCTION

Rapid Evaporative Ionization Mass Spectrometry (REIMS) is a relatively new technology

that is emerging in many areas of science, including human medicine and biological sciences.

REIMS-based tissue analysis generally takes only a few seconds and can provide histological

tissue identification with 90 to 98% correct classification performance (Balog 2013). Recently,

utilization of REIMS in meat products provided very promising results across various

classification scenarios (Balog et al., 2016; Verplanken et al., 2017). Using time-of-flight (TOF)

mass spectrometry, REIMS profiling provides *in situ*, real-time molecularly-resolved

information by ionizing biological samples in real-time without any sample preparation. Waters

Corporation (Wilmslow, UK) has developed this technology and coupled it to a hand-held iKnife

sampling device, allowing for tremendous mobility in the sampling procedure. This technology

would allow for meat quality attributes, such as flavor profile and tenderness, to be predicted and

characterized in real-time via broad biochemical profiling of tissue samples. Unlike other

metabolomic approaches that require tedious sample preparation and analysis times, this

technology could be further developed as an on-line system in the processing environment to

enable meaningful sorting of beef products into categories reflecting tangible differences in

eating characteristics.

Current beef quality grading standards are applied via visual assessment of carcass traits

marbling score, physiological maturity, sex class, and lean texture/firmness. Research has shown

that these grading standards generally separate carcasses based on predicted eating experiences

(Smith et al., 1987; Platter et al., 2003; Emerson et al., 2013). In 2016, only 1.8% of graded beef

carcasses had an overall USDA maturity score of C or greater (Boykin et al., 2017). This indicates that, among carcasses derived from the fed cattle supply, carcass maturity plays a very minimal role in determining quality grades in today's industry and that marbling score is the primary determinant of USDA quality grade. It is the general consensus that as marbling score increases, the probability of a positive eating experience also increases (Emerson et al., 2013). Although marbling is a major component of the grading system, it has shown to account for as little as 5% of variation in eating quality (Wheeler et al., 1994), clearly leaving significant sources of variation unaccounted for during the grading process. Biochemical components of beef muscle are known to influence beef eating quality and may explain variation not accounted for by marbling score alone (Mottram, 1998), but cannot be visually assessed by a human grader or grading camera. Therefore, the objective of experiment 1 was to evaluate the ability of rapid evaporative ionization mass spectrometry to predict various components of beef quality including: carcass type, sensory attributes, and objective tenderness measurements.

Tenderness is one of the most important attributes when determining consumer acceptability of beef (O'Quinn et al., 2012), which was shown to be influenced by cooking method (Yancey et al., 2011). Therefore, it is critical to establish cooking parameters that maximize eating satisfaction, without sacrificing efficiency and practicality of the cooking process. In previous tenderness studies, researchers credited the addition of humidity to the cooking environment as a way to improve the process of tenderization (Kolle et al., 2004; Bowers et al., 2012). Moisture has shown to be useful in the breakdown of protein and the solubilization of collagen, which is especially beneficial when cooking tougher muscles (Cover and Smith, 1956). Collagen shrinks and denatures around 65°C, contributing to the toughening of meat during cooking; however, if held above 70°C for extended periods, denatured collagen

2

will begin to gelatinize and increase tenderness (Purslow, 2005, Bailer and Light, 1989). For this reason, rate of cooking plays a significant role in the tenderness of cooked beef. The objective experiment 2 was to evaluate the influence of relative humidity and oven temperature on external and internal color appearance, protein denaturation, collagen content, shear force values, and sensory attributes of beef strip steaks cooked using varying oven temperatures and relative humidity levels.

CHAPTER II

REVIEW OF LITERATURE – PART I

Beef Grading

The USDA's voluntary beef grading service began in 1926 in an effort for packers to effectively segregate beef carcasses based on inherent quality differences (USDA, 2017). Since the implementation of the grading system, standards have been amended several times throughout the years as we have increased our understanding of the factors influencing beef quality. Until 1989, it was required that a graded carcass receive both a quality and a yield grade; however, the standards were amended so these 2 grades could be applied separately or together. USDA quality grades were established using carcass characteristics to predict eating quality and an overall eating experience. In today's standards, carcasses can qualify for one of eight quality grades: Prime, Choice, Select, Standard, Commercial, Utility, Cutter and Canner. Only steer, heifer, cow, and bullock carcasses qualify for quality grades; whereas, bulls are only eligible for yield grades. USDA quality grading standards are applied via visual assessment of carcass traits marbling score, physiological maturity, sex class, and lean texture/firmness. Marbling score is a visual assessment of the amount of intramuscular fat within the exposed *longissimus* muscle between the 12th and 13th ribs. An overall maturity score is determined by balancing a skeletal maturity score along with a lean maturity score taken from the 12th and 13th rib juncture. Beginning in 2018, the grading standards were amended to allow for dentition to be an optional determination of a carcass being over or under 30 months of age for quality grading purposes, regardless of physiological maturity (USDA, 2017). Once a carcass has been evaluated for each

of the USDA quality grading parameters, it can receive an overall quality grade based on the combination of all attributes.

In the early days of beef grading, grades were applied via the visual assessment of carcass characteristics using trained human graders employed by USDA's Agriculture Marketing Services. A significant portion of beef quality grades are still assigned by human graders; however, beginning in the early 2000s, grading instruments were developed and verified for use in applying official USDA quality and yield grades. Two grading instruments were approved in 2001 to assess ribeye area, official USDA quality grades were approved to be applied via instrumentation in 2007, and two instruments to assess marbling score were approved in 2009. By assigning grades using more objective measurements, consistency is greatly improved, and producers selling livestock based on carcass grades can feel more confident in the accurate assignment of those grades.

Fed beef carcasses are marketed using combinations of both USDA yield and quality grading carcass characteristics, receiving premiums or discounts based on the combination of these characteristics, along with other factors. By applying premiums for both high quality and low yielding carcasses, it provides economic incentives for producers to manage cattle in a way that improves the overall beef supply and increases the consistency in beef products reaching consumers. Starting with Certified Angus Beef in 1978, branded beef programs allow companies to further segregate beef carcasses meeting a unique set of specifications that consider attributes beyond those evaluated by the USDA quality grading system. Each branded program has a unique set of specifications that include attributes to guarantee quality, yield, muscle dimension, breed-type, and production background, among others. The success of Certified Angus Beef ignited the fire for the spread of branded beef programs in the United States. Today, the USDA

certifies 90 individual branded beef programs, which does not include those programs monitored by individual companies and retailers (USDA, 2018). With earned trust from consumers, beef from branded programs can grow to garner premiums beyond those that would be achievable from the USDA grading system alone.

Inadequacies of the Current Beef Grading System

Quality grade is used to predict overall eating quality of beef carcasses as assessed by the combined effects of tenderness, juiciness, and flavor. Generally, as marbling score increases, tenderness, juiciness, and flavor also increase (Platter et al., 2003). Even before the implementation of instrument grading, the USDA quality grading system was effectively segregating carcasses by overall eating quality (Smith et al., 1987). After the implementation of instrument grading, it was further validated that instrument assigned marbling scores continued to segregate carcasses into groups with increased probabilities of a positive overall eating experience (Emerson et al., 2013). Although marbling score is a principal component of the quality grading system, marbling score itself does not explain the entirety of variation in beef sensory attributes (Wheeler et al., 1994; Platter et al., 2003). Both Wheeler et al. (1994) and Platter et al. (2003) found marbling score to explain roughly 5% of the variation in *longissimus* eating quality attributes. On the other hand, O'Quinn et al. (2018) and Emerson et al. (2013) determined marbling score to explain a greater amount of variation in eating quality attributes (14-16% and 61%, respectively). Nevertheless, each of these studies still leaves portions of variation in eating quality left unexplained. Emerson et al. (2013) trained sensory panelists to rate samples for an overall sensory experience based on a combination of individual tenderness, juiciness, and flavor attributes and specifically instructed to not include personal preference. The use of a trained sensory panel to determine an overall sensory experience may partially explain

why the authors found marbling score to account for a greater amount of variation in comparison to other studies.

According to the 2016 National Beef Quality Audit, only 1.8% of graded beef carcasses from fed cattle had overall maturity scores of C or greater (Boykin et al., 2017). This audit occurred before USDA's amendment to their maturity determining standards; therefore, it would be expected that current numbers of graded beef carcasses of fed cattle falling into a C or greater maturity score would be lower today. This is not to say that mature carcasses are not entering packing facilities, but rather, that current USDA quality grading standards do not accurately reflect the merchandising value associated with market cow carcasses (Woerner, 2010). Thus, the majority of mature carcasses do not receive official USDA quality grades. As a result, when only considering the population of beef carcasses receiving USDA quality grades, maturity plays a very minimal role in quality grade determination, leaving significant contribution of final USDA quality grade determination on marbling score.

Higher quality grades do increase the probability of a positive eating experience when consuming beef (Smith et al., 2008; Emerson et al., 2013; O'Quinn et al., 2018). Current grading standards appropriately predict the probability of an overall eating experience. But, it is the variation in individual sensory responses within a quality grade that can be highly variable, particularly within lower grades. Smith et al. (2008) compiled sensory results from 14 previous studies to evaluate the probability of an unsatisfactory eating experience within each quality grade. They found the probability of an unsatisfactory eating experience to be 1 in 33 for Prime, 1 in 10 for Upper 2/3 Choice, 1 in 6 for Low Choice, 1 in 4 for Select, and 1 in 2 for Standard. In support of these findings, similar probabilities have been obtained from others (Tatum, 2015; O'Quinn et al., 2018). Especially within lower quality grades, there is clear variation that is not

7

being accounted for within the current grading system. Discovering the ability to identify beef from lower grading carcasses that will result in a positive eating experience during the grading process would result in added value to currently discounted product. Alternatively, the ability to identify and remove low performing beef carcasses from Prime and Top Choice quality grades would further increase the guarantee the probability of a positive eating experience, allowing packers to obtain additional premiums.

It is well understood that numerous attributes, in addition to marbling score and maturity, significantly influence beef eating quality. These include characteristics such as breed type, muscle fiber type, enzymatic activity, pH, collagen content, production background, fatty acids, amino acids, reducing sugars, and metabolic rates, to name a few (Wheeler et al., 1994; Chriki et al., 2013; Kerth and Miller, 2015; Grayson et al., 2016; O'Quinn et al., 2016; Starkey et al., 2017). This is clearly not an exhaustive list but begins to describe the complexity of sensory development. Some of these attributes can be estimated by visual assessment; i.e., hide color or neck hump height, and have been incorporated into various branded beef programs. Nevertheless, the majority of these attributes are biochemical components that cannot be visually assessed. Others could be verified via certification; however, this can be logistically cumbersome on a large-scale application. Therefore, an instrument with the ability to assess biochemical components of fat and lean at chain speed would prove to be most accurate in not only segregating carcasses into meaningful eating quality groups, but also verifying other claims such as breed-type and production background. Coupled with the current USDA grading standards, an instrument of this caliber would prove to have substantial monetary value for the beef industry.

Beef Grading Instruments

Although the first beef grading instrument was not approved for use until 2001, the USDA identified the importance and necessity to develop instrument grading systems in 1978 (Woerner and Belk, 2008). In conjunction with the National Aeronautics and Space Administration (NASA) in 1979, the USDA's Food Safety and Quality Service (FSQS) recognized ultrasound and video image analysis (VIA) as two technologies with potential for the assessment of beef quality and yield grading characteristics (Cross and Whittaker, 1992). At the time, it was suggested that the precision and consistency of an objective grading system would improve the specificity of the grading system and would benefit producers, packers, and consumers. As a result, research progressed with the further evaluation of VIA as an instrument grading technique (Cross and Whittaker, 1992). Early work showed that beef carcass yield prediction using VIA measurements increased (93.6% vs. 84.42%) the coefficient of determination of equations when compared to yields predicted using non-instrument measurements (Cross et al., 1983). The VIA system utilized chilled and ribbed carcass; however, in 1984, industry leaders decided to shift the instrument grading focus away from chilled and ribbed carcasses towards unchilled and unribbed carcasses (Woerner and Belk, 2008). Thus, instrument focus shifted away from VIA towards ultrasound analysis to predict ribeye area, fat thickness, and marbling score. Little progress was made in the development of ultrasound as an online beef grading tool; thus, instrument grading focus was once again placed on VIA (Cross and Whittaker, 1992).

In order for a grading instrument to be successful, it not only needs to accurately measure predictors, but it needs to be rapid enough to handle chain speed and be robust enough to handle the extreme environmental conditions of carcass coolers. In 1990, the National Cattleman's

Association created an Instrument Grading Subcommittee charged with the task of determining the most promising avenues for instrument grading in the beef industry. An initial outcome of this committee was to establishment a set of parameters that must be met for the successful installment of an instrument grading system. They identified that an instrument grading system must 1) predict the percentage of lean, marbling, and skeletal maturity with high accuracy, 2) it must produce repeatable measurements of individual factors, 3) it must be completely automated, including the interpretation of the image or the output, 4) it must be able to predict all necessary carcass traits at a rate that can be maintained with production speeds, 5) it must be able to withstand extreme changes in temperature (0-40°C) and humidity (up to 100%), 6) it must be tamper proof to prevent assessment errors, and 7) recalibration must be precise, quick, and easy (Cross and Whittaker, 1992).

Through continued research, VIA has shown success in accurately predicting beef yield characteristics (Cross et al., 1983; Wassenberg et al., 1986; Shackelford et al., 1998; Lorenzo et al., 2018). However, error still occurred, primarily due to inadequacies in fat estimations. Belk et al. (1998) determined that VIA could accurately measure preliminary yield grade (PYG) and ribeye area (REA) but could not appropriately evaluate the more subjective measurement of adjusted preliminary yield grade (APYG) to account for total carcass fatness and/or dressing defects. Thus, it was suggested that VIA yield grade assessment augmented with USDA grader adjusted fat thickness provided a more accurate and efficient use of VIA determined yield grades. Further development of VIA in the prediction of APYG resulted in improved accuracy and the ability for image analysis variables to account for 88% of the variation between calibration and predicted measurements for APYG (Shackelford et al., 2003). Video image analysis was approved for measuring ribeye in 2001, approved to calculate USDA yield grade in

2005, and began official use for applying yield grades in 2007 (Mafi et al., 2014). While VIA proved its ability to calculate yield parameters, it is has shown to be less successful in assessing marbling score. Shackelford et al. (2003) showed VIA assessment of marbling score to account for 76% of variation, concluding that it did not meet the criteria for industry application.

Advancements in VIA technology led to 2 cameras with the ability to accurately assess marbling score: CVS (Computer Vision System; RMS Research Management Systems, USA, Inc., Fort Collins, CO) and VBG2000 (E + V Technology, Oranienburg, Germany). In 2006, each of these systems managed to surpass the first phase of USDA's Performance Requirements for Instrument Marbling Evaluation (PRIME) program (Woerner and Belk, 2008). The USDA Agriculture Marketing Service Livestock, Poultry, and Seed Program (USDA-AMS LPS Program) developed PRIME to provide performance standards for instruments used to assess beef marbling scores (USDA-AMS LPS, 2006). To pass the first phase (PRIME I), an instrument must demonstrate its ability to repeatedly predict marbling score of stationary carcasses. To do so, marbling score must be measured 3 times per carcass and the values of each of the 3 measurements must be within 20 marbling score units of the average. The second phase (PRIME II) evaluates the accuracy and precision of the instrument at production speeds compared to marbling scores assigned by a panel of 5 expert human evaluators. In order to pass PRIME II and receive final approval for quality grade assignment, the instrument must have an average residual of 0 ± 10 marbling score units compared to panel assigned score, a standard deviation of residuals ≤ 35 marbling score units, and line of best fit with a slope of 0.000 ± 0.075 when plotting the residuals from panel assigned marbling score versus the instrument marbling score (USDA-AMS LPS, 2006). The CVS system proved to have an accuracy of 89% with repeatability greater than 99.5% in determining marbling score (Moore et al., 2010).

Additionally, the E + V system appropriately applies marbling scores to segregate beef carcasses into meaningful sensory-based groups (Emerson et al., 2013).

Tenderness Prediction Instruments

Methods to either identify or predict beef tenderness during carcass merchandising has been an area of great interest. Consumer studies have identified a willingness to pay premiums for beef that can be guaranteed tender (Shackelford et al., 2001); however, the current USDA quality grading system does not completely segregate between tough and tender beef, particularly at lower grades (O'Quinn et al., 2018). With lower quality carcasses either not qualifying for premiums or receiving discounts, value is being lost from tender beef within these quality grades, as a greater probability of being tough is assumed due to a lack of intramuscular fat. On the contrary, carcasses meeting specifications for higher quality grades, especially those further qualifying for premium boxed beef programs, do not completely remove tough beef from those populations. Branded beef programs rely on their reputation and the consistency of high-quality beef in order to gain consumer trust, earn repeat business, and garner significant premiums for their products. Although the likelihood is often reduced, inclusion of a low percentage of tough beef into these programs could result in lost business. Therefore, the ability to identify tough versus tender beef prior to carcass merchandising would provide increased marketability, consumer trust, and added value throughout the entire beef system.

Warner-Bratzler shear force (WBS) has long been an industry standard as an objective prediction of beef tenderness. However, the protocol is relatively time consuming and results in the destruction of product. Shackelford et al. (1999) developed and validated a rapid alternative to WBS, called slice shear force (SSF). The new method proved to be repeatable ($r = 0.89$) and have the ability to accurately segregate samples into tender, intermediate, and tough

classifications with an overall accuracy of 94.4% (Shackelford et al., 1999). Although proving to be more rapid and have a simpler protocol than its WBS predecessor, SSF is not able to keep up with chain speed and still requires an entire 2.54 cm think strip steak for analysis. For these reasons, producers believe this method too costly as verification for guaranteed tender programs (Wheeler et al., 2002). Regardless, the SSF method has retained great popularity within academic research, as well as, with individual companies as an off-line method to track tenderness of branded beef programs (Woerner and Belk, 2008).

In addition to SSF, less destructive approaches have been evaluated for on-line tenderness prediction, but an instrument method to predict beef tenderness has yet to be implemented into production systems. Evaluated methods include: Tendertec Tenderness Probe, objective color measurements, near-infrared (NIR) spectroscopy, and hyperspectral images, among others. The TenderTec Tenderness Probe uses an electromechanical penetrometer inserted into the *longissimus* muscle. It was originally identified by the Australian Meat Research Corporation for its potential ability to predict beef tenderness; however, when used on US beef carcasses, it was found to only have a slight tenderness predictive ability with mature, but not youthful carcasses (Belk et al., 2001). Objective b* measurements of *longissimus* muscle have been found to have a stronger relationship to sensory tenderness than marbling score, but still only explained 14% of tenderness variation (Wulf et al., 1997). Later integration of objective lean and fat color measurements with predicted marbling and adjusted REA showed improvement in prediction accuracy to using b* values alone (Vote et al., 2003). In this study, all measurements were obtained with a CVS camera adapted with a BeefCam™ module. Although the BeefCam™ system was able to correctly identify tender carcasses with an overall accuracy of 80%, this accuracy was significantly affected when samples were variable in marbling score.

Thus, it was concluded that the BeefCam™ system would be most effective when implemented after the administration of an USDA quality grade (Vote et al., 2003).

Tenderness prediction using NIR-spectroscopy was first evaluated by Misumoto et al. (1991), achieving 68% accuracy in predicting beef tenderness. Near-infrared spectroscopy is the measurement of the absorbance of electromagnetic radiation, which can provide information regarding the biochemical makeup of a substance. Later work found NIR to explain 67% of the variation in shear force values, but was able to segregate between tough and tender samples with an overall accuracy of 79% (Park et al., 1998). More recently, hyperspectral imaging (HSI) has been evaluated for its use in predicting beef tenderness. Hyperspectral imaging is essentially a combination of NIR and VIA, having the ability to capture textural information from VIA and molecular information from NIR (Konda et al., 2008). By separating beef *longissimus* samples into tough and tender categories based on SSF values, HSI predicted the classification of tender and tough samples with 96.3% and 62.5%, respectively, for an overall accuracy of 77% (Konda et al., 2008).

Of any of the technologies discussed thus far, prediction accuracies presented by Konda et al. (2008) captured from HSI clearly provided the greatest prediction accuracy. However, there was a significant class imbalance issue with this data set, as only 5% of collected samples fell into the tough category. Therefore, even without prediction using HSI, there is still 95% chance of selecting a tender sample if one was chosen at random. Additionally, misclassification rates were determined using cross-validation of a single data set, as opposed to building the model on a train set and testing the accuracy of the model on a new test data set. Because models are built to specifically fit the training set, training error rates are typically overly optimistic and are not necessarily an indication of an appropriate model (Ghatak, 2017). Furthermore, the authors used

a canonical discriminant model with partial least squares regression images as the predictor variables. Both canonical discriminant analysis and partial least squares regression are supervised statistical methods, meaning that consideration for the response is taken to discover the variation in the data set. Combining both of these modeling techniques, selection for variation specifically explaining the tenderness response would have been performed twice. Especially without using separate training and test data sets, the combination of two supervised statistical methods would greatly increase the risk of overfitting the prediction model. Later work using the same HSI technology, but designating training and test sets and having a slightly greater proportion of tough samples (18%), reported an overall accuracy of 59.2% (Konda Naganathan et al., 2015), supporting the criticisms above. Even with a low overall prediction accuracy, the authors were still able to report a tenderness certification accuracy of 87.6%. Out of the samples predicted as tender, tenderness certification accuracy was calculated as the percent of samples that were truly tender. Although this is arguably a meaningful metric for industry practice, it does not reflect the fact that most models still failed to predict over 50% of truly tough samples as tough. Again, the lack of class balance is not appropriately reflected in this calculation.

Rapid Evaporative Ionization Mass Spectrometry (REIMS)

Rapid Evaporative Ionization Mass Spectrometry (REIMS) is a relatively new form of mass spectrometry originally developed for the medical field to identify cancerous tissue in real-time during removal surgeries. However, its abilities have been recognized for use in other industries, particularly to assess food quality and authenticity. The uniqueness of REIMS, compared to other mass spectrometry methods, comes from its ability to quickly extract and ionize molecules with a handheld device, requiring absolutely zero sample preparation (Waters

Corporation, 2015). The current collection device is a handheld surgical knife (termed iKnife) attached by a 2-3 m tubing, allowing for incredible convenience and mobility during sampling. When attached to a time-of-flight (TOF) analyzer, the operator is provided with a highly accurate mass spectrum in a matter of seconds. Using mass spectra from known test samples, models can then be developed to predict and classify the tissue of interest (Balog et al., 2010). Recent application of REIMS has slowly moved into the meat industry with interests in species authentication, eating quality prediction, and residue testing (Balog et al., 2015; Verplanken et al., 2017; Guitton et al., 2018).

The uniqueness of REIMS does not come from methods used to detect ions, but rather in the sample collection and ionization steps. Few metabolomic techniques allow for a complete lack of sample preparation. Some sample preparation methods are more complicated than others, but regardless, sample preparation not only requires time, but it also introduces a greater risk of technical error. Currently, the source of collection and ionization is a handheld surgical knife with a metal tip and the sample must be placed on a return electrode mat. When the iKnife contacts the surface of the sample, it creates an electric current that heats the metal tip, cauterizing the sample. This creates an aerosol of gas-phased clusters of both ionized and neutral molecules containing unique components of the tissue. The knife is connected to a vacuum tube that draws the aerosol into the machine through a transfer capillary. The stream of clustered ions and neutral molecules then reaches a heated impactor that disrupts the cluster and ionizes the remaining neutral molecules (Golf et al., 2015). The molecules are pushed into a StepWave ion guide to remove gas and other neutral contaminants, significantly increasing sensitivity during the detection phase. The StepWave system is unique in that it is an off-axis guide that pushes ions up into the analyzer and gas/contaminants down out of the machine (Waters Corporation,

2015). Remaining ions are then detected with a TOF analyzer. The ionization source is a soft ionization method; therefore, most ions will be adducted ions without fragmentation. The resulting mass spectra can then go through chemometric profiling, similar to methods used with other mass spectrometry data.

Prediction of various meat attributes using mass spectra collected with REIMS has shown incredible potential for its application in various scenarios. As of now, REIMS mass spectra have most commonly been analyzed using principal component analysis alone, or as a dimension reduction technique coupled with linear discriminant analysis. Balog et al. (2016) showed that REIMS had the ability to differentiate between various mammalian meat species and beef breeds with 100% and 97% accuracy, respectively. Similarly, REIMS has demonstrated a nearly perfect prediction rate (98.99%) in the identification of several fish species (Black et al., 2017). Furthermore, Guitton et al. (2018) successfully identified several porcine muscles from animals fed with accuracies greater than 95%. Verplanken et al. (2017) even successfully segregated pork carcasses with and without boar taint, indicating potential of REIMS in meat eating quality prediction. With rapid analysis, high specificity, and lack of sample preparation, REIMS could have significant implications in the prediction and verification of several beef quality characteristics.

Currently, identified compounds from REIMS output have almost been entirely restricted to various lipid components (Balog et al., 2016). Phospholipids are major components of cell membranes, thus have allowed for successful determination of histological characteristics of tissues (St John et al., 2017). From a meat quality standpoint, lipid profiles greatly impact flavor development, with certain lipids associating with desirable flavor attributes, while others associate with severe off-flavor development (Kerth and Miller, 2015). Establishing the

17

capabilities of REIMS to identify other compounds, such as amino acid profiles and proteins, would add further value in its use in meat quality research. Although much of the current prediction capabilities of REIMS has have been based on lipid profiles, it has shown to be able to successfully identify volatile compounds, flavonoids, and carbohydrates from honey samples (Stead, 2016). Further understanding the ability of REIMS to identify other biochemical components would significantly increase its applications.

Predictive Modeling

With advancements in technology and the ability to track, collect, and store enormous amounts of data, prediction modeling has become exceedingly popular in numerous applications, from e-mail spam filters to identifying credit card fraud. Predictive modeling can be defined as "the process of developing a mathematical tool or model that generates an accurate prediction" (Kuhn and Johnson, 2016). Predictive modeling is not concerned with understanding why an outcome happens, but rather identifying the probability of an outcome occurring. Interpretability of a prediction model is many times a secondary objective but can be a cumbersome task. As data sets become larger and pressure is placed on providing an accurate prediction of an outcome, models tend to become more complex and more difficult to interpret (Kuhn and Johnson, 2016). Some predictive models are more flexible, whereas others are more restrictive. Linear models are considered restrictive because they require a linear relationship between responses and predictors, resulting in a relatively interpretable relationship. Flexible models, on the other hand, do not rely on a linear relationship. Instead, the relationship between the response and predictor can take various shapes. But, the estimation of the relationship between response and predictor can become incredibly complex and difficult to interpret (James et al., 2013). Whether or not focus is placed on prediction or interpretation is largely relative to the issue or

severity of the outcome at hand. Therefore, the primary objective of prediction or interpretation needs to be decided upon by investigators in order to appropriately address the question of concern.

Predictive models can be developed to predict either quantitative or qualitative outcomes. Although not always the case, predictive models utilizing quantitative responses are considered regression problems, whereas, models utilizing qualitative responses are considered classification problems (James et al., 2013). Responses of regression problems are required to be either continuous values or ordered numerical values and the prediction is of a numerical output value. Alternatively, classification responses are unordered categories, thus, the prediction assigns the output into a class. With quantitative outcomes usually being more specific and numerous than qualitative outcomes, numeric response variables can be grouped into categorical classes that provide meaningful separation in a response. By creating a response that is less specific, an increase in prediction accuracy would be expected. For instance, one may find better accuracy in predicting a USDA beef quality grade as opposed to predicting a specific marbling score. The choice of predictive model is based on the characteristics of the response variable and whether it is numerical or categorical (James et al., 2013). The majority of predictive models are capable of handling both numeric and categorical predictors.

High-Dimensional Data

Advancements in bioanalytic techniques now allow researchers to collect highly abundant amounts of data in a relatively cost-effective and timely manner. This has led to the development of various "omics" fields within biological sciences, which aim at understanding the entirety of a biological system and how it elicits different outcomes. The various omics fields are linked together in a cascade-like manner in descending order of: genomics, transcriptomics,

proteomics, and metabolomics (Dettmer et al., 2007). Each omics field is related to the next, with metabolomics having the most direct relationship with an observed phenotype. With the objective of omics approaches to evaluate the entire system, the resulting outputs are high-dimension data sets. A high-dimensional data set can be described as one that has more variables than it does observations (James et al., 2013). Although high-dimensional data contains useful information, it also contains too many irrelevant features, which makes fitting a model difficult, often referred to as the "curse of dimensionality." Irrelevant features are known as noise, which is information that cannot be captured, but is a source of error in predictive models (Ghatak, 2017). For that reason, it is critical to apply methods to reduce dimensionality of the data and identify features that provide relevant information about the response of interest.

Common dimension reduction techniques for metabolomics data include Principal Component Analysis (PCA) and Partial Least Squares (PLS; Maitra and Yan, 2008). Both of these methods reduce the dimension of a data set by finding linear combinations that best explain the variation within the original data to create a new set of latent predictor variables (James et al., 2013). The newly predicted latent variables for each observation for each principal component are referred to as factor scores (Abdi and Williams, 2010). Factor scores can be plotted either 2 or 3-dimensionally to visually evaluate the spatial projections onto the principal components, which is useful in determining clusters of similar samples or potential outliers. Loadings, another common output of PCA and PLS, describe the correlation between the original variables with each component, providing information on the weight each variable had in calculating factor scores (Abdi and Williams, 2010). Loadings are commonly plotted between 2 components, allowing for a visual representation of the coefficients assigned to each variable. Additionally, imposing loading onto scores plots provide further visualization and interpretability of how

individual variables drive the projection of individual sample observations on the principal components.

Although the 2 methods share several similarities, a major difference is that PCA is considered an unsupervised method, whereas, PLS is considered a supervised method (Maitra and Yan, 2008). Principal component analysis does not consider the response variable when extracting sources of variation (Kuhn and Johnson, 2016). It proceeds "unguided"; thus, considered unsupervised. In some situations, this can be troublesome because the greatest sources of variation within a data set may not necessarily be related to the response. The first principal component will always explain the greatest amount of variation, the second principal component will explain the second greatest amount of variation, and so forth. Additionally, each component will be uncorrelated to the others, resolving issues with highly-collinear variables that is inevitable with high-dimension data sets. Principal components can be a predecessor for Principal Component Regression, where factor scores can be regressed against one or more dependent variables (Geladi and Kowalski, 1986). Because it reduces dimension and removes collinearity, PCA is commonly used to preprocess data before insertion into other regression or classification models.

Several attributes of PCA hold true with PLS models. In contrast to PCA, however, PLS uses the response to find variation in the predictors that best explain the response (Garthwaite, 1994). Because it is finding the linear combinations of predictor variables that best explain the response, the first component will not necessarily explain the greatest amount of variation within the predictor variables like with PCA. A PLS model will simultaneously evaluate three objectives during the model fitting process: 1) the best explanation of the X space, 2) the best explanation of the Y space, and 3) the greatest relationship between the X and Y spaces (Ghatak,

2017). Therefore, PLS will calculate two sets of scores that best maximize the covariance between the X and Y-spaces (Ghatak, 2017). Even if PCA and PLS create models with similar predictive abilities, PLS will typically do so using fewer components and a simpler model. As a general rule of thumb, simpler models are preferred over more complicated models (Bro and Smilde, 2003). When models become increasingly more complex with more variables or more components, slight changes in values of new data have a greater probability of resulting in prediction error. For these reasons, PLS is typically preferred over PCA. Furthermore, PLS is better suited to handle data sets that have more variables than observations, leading to an increased popularity of its use in chemometric analysis (Höskuldsson, 1988).

Data Pre-Processing

Data pre-processing is a step in predictive modeling that can severely alter model accuracy (Kuhn and Johnson, 2016). The effectiveness of pre-processing methods depends on the characteristics of the data set and their relationship with the response and there is no clear answer for selection of pre-processing methods (Van Den Berg et al., 2006). Therefore, the decision of which pre-processing methods to use will frequently be at the discretion of the researcher, their knowledge of the relationship between dependent and independent variables, or simply choosing methods that reduce prediction error. Particularly with large data sets, PLS or PCA methods described above can fall into the category of pre-processing techniques if they are used to reduce collinearity and dimensionality before utilization in further predictive models. This type of pre-processing would result in the utilization of newly calculated latent variables as opposed to the originally collected predictor variables. However, other pre-processing methods preserve the integrity of the original predictors and rely on transformations of observed values. If a robust model can be built without latent variables, the interpretability of the model would likely

increase. Of the methods that rely on data transformations of individual predictors, mean-centering, scaling, and skewness transformations or common choices in chemometric data analysis (Kuhn and Johnson, 2016).

To mean-center data, the overall mean is subtracted from each individual observation so that the variable has a mean of zero (Van Den Berg et al., 2006). This adjusts for the offset between variables that are found in relatively high abundances and those that are found in relatively low abundances, converting the data from an interval scale to a ratio-scale (Bro and Smilde, 2003). Scaling refers to dividing each variable by a unique scaling factor so that each variable in the matrix is represented on a similar scale (Van Den Berg et al., 2006). Unit-variance (UV) scaling is a commonly applied scaling method to metabolomics data, in which each observation is divided by the standard deviation. By standardizing variable scale using standard deviation, emphasis is placed on variables with greater amounts of variation regardless of how numerically small or large the original observations may be. This is especially beneficial in scenarios where included variables are not measured using the same scale or when there are large fold differences between metabolites.

Linear Discriminant Analysis

Linear discriminant analysis (LDA) is an approach commonly used in classification problems of linear data. It is similar to PLS in that it is a supervised method that finds components that best maximize variance, while also maximizing separation between classes. Linear discriminant analysis calculates linear discriminants that apply weights to each variable so that individual observations can be converted and projected into a 2 or 3-dimensional space as scores. The maximum number of linear discriminants is equal to 1 minus the number of classification categories. Additionally, discriminant scores are calculated to identify how well

23

each function predicts classification and the sum will always equal 100% of explained variation. However, unlike PLS, LDA is not appropriate to handle high-dimension data sets and requires that the number of predictors be less than the number of observations (Zelterman, 2015). As a general rule of thumb, the number of observations needs to be greater than 5 times the number of predictors. For this reason, LDA is not suited to handle raw omics data, although it has shown to have high classification accuracy when coupled with a dimension reduction technique (Balog et al., 2016; St John et al., 2017; Guitton et al., 2018). As mentioned above, PCA and PLS have proven to be appropriate methods for reducing dimension and collinearity of large data sets and are commonly coupled with LDA analysis.

LITERATURE CITED

Abdi, H., and L. J. Williams. 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2:433–459.

Balog, J., S. Kumar, J. Alexander, O. Golf, J. Huang, T. Wiggins, N. Abbassi-Ghadi, A. Enyedi, S. Kacska, J. Kinross, G. B. Hanna, J. K. Nicholson, and Z. Takats. 2015. In Vivo Endoscopic Tissue Identification by Rapid Evaporative Ionization Mass Spectrometry (REIMS). Angew. Chemie - Int. Ed. 54:11059–11062.

Balog, J., D. Perenyi, C. Guallar-Hoyas, A. Egri, S. D. Pringle, S. Stead, O. P. Chevallier, C. T. Elliott, and Z. Takats. 2016. Identification of the Species of Origin for Meat Products by Rapid Evaporative Ionization Mass Spectrometry. J. Agric. Food Chem. 64:4793–4800.

Balog, J., T. Szaniszlo, K. Schaefer, J. Denes, A. Lopata, L. Godorhazy, D. Szalay, L. Balogh, L. Sasi-szabo, M. Toth, and Z. Takats. 2010. Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry. Anal. Chem. 82:7343–7350.

Belk, K. E., M. H. George, J. D. Tatum, G. G. Hilton, R. K. Miller, M. Koohmaraie, J. O. Reagan, and G. C. Smith. 2001. Evaluation of the Tendertec beef grading instrument to predict the tenderness of steaks from beef carcasses. J. Anim. Sci. 79:688–697.

Van Den Berg, R. A., H. Cj Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. Van Der Werf. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data.

Black, C., Olivier, P. Chevallier, S. A. Haughey, J. Balog, S. Stead, S. D. Pringle, M. V Riina, F. Martucci, P. L. Acutis, M. Morris, D. S. Nikolopoulos, Z. Takats, and C. T. Elliott. 2017. A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry. Metabolomics 13.

Boykin, C. A., L. C. Eastwood, M. K. Harris, D. S. Hale, C. R. Kerth, D. B. Griffin, A. N. Arnold, J. D. Hasty, K. E. Belk, D. R. Woerner, R. J. Delmore, J. N. Martin, D. L. VanOverbeke, G. G. Mafi, M. M. Pfeiffer, T. E. Lawrence, T. J. McEvers, T. B. Schmidt, R. J. Maddock, D. D. Johnson, C. C. Carr, J. M. Scheffler, T. D. Pringle, A. M. Stelzleni, J. Gottlieb, and J. W. Savell. 2017. National Beef Quality Audit–2016: In-plant survey of carcass characteristics related to quality, quantity, and value of fed steers and heifers1. J. Anim. Sci. 95:2993–3002.

Bro, R., and A. K. Smilde. 2003. Centering and scaling in component analysis. J. Chemom. 17:16–33.

Chriki, S., G. Renand, B. Picard, D. Micol, L. Journaux, and J. F. Hocquette. 2013. Meta-analysis of the relationships between beef tenderness and muscle characteristics. Livest. Sci. 155:424–434.

Cross, H. R., D. A. Gilliland, P. R. Durland, and S. Seideman. 1983. Beef Carcass Evaluation by Use of a Video Image Analysis System. J. Anim. Sci. 57:908–917.

Cross, H. R., and A. D. Whittaker. 1992. The Role of Instrument Grading in a Beef Value-Based Marketing System'. J. Anim. Sci. 70:984–989.

Dettmer, K., P. A. Aronov, and B. D. Hammock. 2007. Mass spectrometry-based metabolomics. Mass Spectrom. Rev. 26:51–78.

Emerson, M. R., D. R. Woerner, K. E. Belk, and J. D. Tatum. 2013. Effectiveness of USDA instrument-based marbling measurements for categorizing beef carcasses according to differences in longissimus muscle sensory attributes. J. Anim. Sci. 91:1024–1034.

Garthwaite, P. H. 1994. An lnterpretation of Partial Least Squares. J. Am. Stat. Assoc. 89.

Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185:1–17.

Ghatak, A. 2017. Machine Learning with R. Springer, Singapore.

Golf, O., N. Strittmatter, T. Karancsi, S. D. Pringle, A. V. M. Speller, A. Mroz, J. M. Kinross, N. Abbassi-Ghadi, E. A. Jones, and Z. Takats. 2015. Rapid evaporative ionization mass spectrometry imaging platform for direct mapping from bulk tissue and bacterial growth media. Anal. Chem. 87:2527–2534.

Grayson, A. L., S. D. Shackelford, D. A. King, R. O. McKeith, R. K. Miller, and T. L. Wheeler. 2016. The effects of degree of dark cutting on tenderness and sensory attributes of beef. J. Anim. Sci. 94:2583–2591.

Guitton, Y., G. Dervilly-Pinel, R. Jandova, S. Stead, Z. Takats, and B. Le Bizec. 2018. Rapid evaporative ionisation mass spectrometry and chemometrics for high-throughput screening of growth promoters in meat producing animals. Food Addit. Contam. Part A:1–11. Available from: https://www.tandfonline.com/doi/full/10.1080/19440049.2017.1421778
Höskuldsson, A. 1988. PLS regression methods. J. Chemom. 2:211–228.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning with Applications in R. Springer, New York, NY.

Kerth, C. R., and R. K. Miller. 2015. Beef flavor: A review from chemistry to consumer. J. Sci. Food Agric. 95:2783–2798.

Konda, G., N. Ae, L. M. Grimes, A. E. Jeyamkondan, S. Ae, C. R. Calkins, A. E. Ashok, S. Ae, and G. E. Meyer. 2008. Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. Sens. Instrum. Food Qual.

Konda Naganathan, G., K. Cluff, A. Samal, C. R. Calkins, D. D. Jones, C. L. Lorenzen, and J. Subbiah. 2015. Hyperspectral imaging of ribeye muscle on hanging beef carcasses for tenderness assessment. Comput. Electron. Agric. 116:55–64.

Kuhn, M., and K. Johnson. 2016. Applied Predictive Modeling. Springer, New York, NY.

Lorenzo, J. M., C. M. Guedes, R. Agregán, M. V. Sarriés, D. Franco, and S. R. Silva. 2018. Prediction of foal carcass composition and wholesale cut yields by using video image analysis. animal 12:174–182.

Mafi, G., B. Harsh, and J. Scanga. 2014. Review of Instrument Augmented Assessment of USDA Beef Carcass Quality Grades. In: 67th Annual Reciprocal Meat Conference.

Maitra, S., and J. Yan. 2008. Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression.

Moore, C. B., P. D. Bass, M. D. Green, P. L. Chapman, M. E. O'Connor, L. D. Yates, J. A. Scanga, J. D. Tatum, G. C. Smith, and K. E. Belk. 2010. Establishing an appropriate mode of comparison for measuring the performance of marbling score output from video image analysis beef carcass grading systems1,2. J. Anim. Sci. 88:2464–2475.

O'Quinn, T. G., J. F. Legako, J. C. Brooks, and M. F. Miller. 2018. Evaluation of the contribution of tenderness, juiciness, and flavor to the overall consumer beef eating experience. Transl. Anim. Sci.

O'Quinn, T. G., D. R. Woerner, T. E. Engle, P. L. Chapman, J. F. Legako, J. C. Brooks, K. E. Belk, and J. D. Tatum. 2016. Identifying consumer preferences for specific beef flavor characteristics in relation to cattle production and postmortem processing parameters. Meat Sci. 112:90–102.

Park, B., Y. R. Chen, W. R. Hruschka, S. D. Shackelford, and M. Koohmaraie. 1998. Near-infrared reflectance analysis for predicting beef longissimus tenderness. J. Anim. Sci. 76:2115.

Platter, W. J., J. D. Tatum, K. E. Belk, P. L. Chapman, J. A. Scanga, and G. C. Smith. 2003. Relationships of consumer sensory ratings, marbling score, and shear force value to consumer acceptance of beef strip loin steaks. J. Anim. Sci 81:2741–2750.

Platter, W. J., J. D. Tatum, K. E. Belk, P. L. Chapman, J. A. Scanga, and G. C. Smith. 2003. Relationships of consumer sensory ratings, marbling score, and shear force value to consumer acceptance of beef strip loin steaks. J. Anim. Sci. 81:2741.

Shackelford, S. D., T. L. Wheeler, and M. Koohmaraie. 1998. Coupling of Image Analysis and Tenderness Classification to Simultaneously Evaluate Carcass Cutability, Longissimus Area, Subprimal Cut Weights, and Tenderness of Beef.

Shackelford, S. D., T. L. Wheeler, and M. Koohmaraie. 2003. On-line prediction of yield grade, longissimus muscle area, preliminary yield grade, adjusted preliminary yield grade, and marbling score using the MARC beef carcass image analysis system. J. Anim. Sci. 81:150–155.

Shackelford, S. D., T. L. Wheeler, M. Koohmaraie, and R. L. Hruska. 1999. Tenderness Classification of Beef: II. Design and Analysis of a System to Measure Beef Longissimus Shear Force Under Commercial Processing Conditions 1. J. Anim. Sci 77:1474–1481.

Shackelford, S. D., T. L. Wheeler, M. K. Meade, J. O. Reagan, B. L. Byrnes, and M. Koohmaraie. 2001. Consumer impressions of Tender Select beef. J. Anim. Sci. 79:2605.

Smith, G. C., J. W. Savell, H. R. Cross, Z. L. Carpenter, C. E. Murphey, G. W. Davis, H. C. Abraham, F. C. Parrish, and B. W. Berry. 1987. Relationship of USDA Quality Grades to Palatability of Cooked Beef. J. Food Qual. 10:269–286.

Smith, G. C., J. D. Tatum, and K. E. Belk. 2008. International perspective: characterisation of United States Department of Agriculture and Meat Standards Australia systems for assessing beef quality. Aust. J. Exp. Agric. 48:1465.

St John, E. R., J. Balog, J. S. McKenzie, M. Rossi, A. Covington, L. Muirhead, Z. Bodai, F. Rosini, A. V. M. Speller, S. Shousha, R. Ramakrishnan, A. Darzi, Z. Takats, and D. R. Leff. 2017. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. Breast Cancer Res. 19:59.

Starkey, C. P., G. H. Geesink, R. van de Ven, and D. L. Hopkins. 2017. The relationship between shear force, compression, collagen characteristics, desmin degradation and sarcomere length in lamb biceps femoris. Meat Sci. 126:18–21.

Tatum, J. D. 2015. Recent Trends: Beef Quality, Value and Price.

USDA. 2006. Performance requirements for Instrument Marbling Evaluation (PRIME II). Implementation and verification of operational procedures. Livestock, Poultry, and Seed Program, Agric. Marketing Ser., USDA, Washington, DC.

USDA. 2017. United States Standards for Grades of Carcass Beef. Livestock, Poultry, and Seed Program, Agric. Marketing Ser., USDA, Washington, DC.

USDA. 2018. USDA Certified Beef Programs. Agric. Marketing Ser., USDA, Washington, DC.

Verplanken, K., S. Stead, R. Jandova, C. Van Poucke, J. Claereboudt, J. Vanden Bussche, S. De Saeger, Z. Takats, J. Wauters, and L. Vanhaecke. 2017. Rapid evaporative ionization mass spectrometry for high-throughput screening in food analysis: The case of boar taint. Talanta 169:30–36.

Vote, D. J., K. E. Belk, J. D. Tatum, J. A. Scanga, and G. C. Smith. 2003. Online prediction of beef tenderness using a computer vision system equipped with a BeefCam module. J. Anim. Sci. 81:457–465.

Wassenberg, R. L., D. M. Allen, and K. E. Kemp. 1986. Video Image Analysis Prediction of Total Kilograms and Percent Primal Lean and Fat Yield of Beef Carcasses. J. Anim. Sci. 62:1609–1616.

Wheeler, T. L., L. V. Cundiff, and R. M. Koch. 1994. Effect of marbling degree on beef palatability in Bos taurus and Bos indicus cattle. J. Anim. Sci. 72:3145–3151.

Wheeler, T. L., D. Vote, J. M. Leheska, S. D. Shackelford, and K. E. Belk. 2002. The Efficacy of Three Objective Systems for Identifying Beef Cuts That Can Be Guaranteed Tender.

Woerner, and K. E. Belk. 2008. The History of Instrument Assessment of Beef. Report Submitted to the National Cattlemen's Beef Association, Denver, CO.

Woerner, D. R. 2010. Beef from Market Cows. Report Submitted to the National Cattlemen's Beef Association, Denver, CO.

Wulf, D. M., S. F. O'Connor, J. D. Tatum, and G. C. Smith. 1997. Using objective measures of muscle color to predict beef longissimus tenderness. J. Anim. Sci. 75:684-692.

Zelterman, D. 2015. Applied Multivariate Statistics with R. In: Spinger International Publishing, Switzerland.

CHAPTER III

ASSESSMENT OF EVAPORATIVE IONIZATION MASS SPECTROMETRY (REIMS) TO

CHARACTERIZE BEEF

Introduction

Rapid Evaporative Ionization Mass Spectrometry (REIMS) is a relatively new technology

that is emerging in many areas of science, including human medicine and biological sciences.

REIMS-based tissue analysis generally takes only a few seconds and can provide histological

tissue identification with 90−98% correct classification performance (Balog 2013). Recently,

utilization of REIMS in meat products provided very promising results across various

classification scenarios (Balog et al., 2016; Verplanken et al., 2017). Using time-of-flight (TOF)

mass spectrometry, REIMS profiling provides *in situ*, real-time molecularly-resolved

information by ionizing biological samples in real-time without any sample preparation. Waters

Corporation (Wilmslow, UK) has developed this technology coupled to a hand-held iKnife

sampling device, allowing for tremendous mobility in the sampling procedure. For the first time,

this technology would allow for meat quality attributes, such as flavor profile and tenderness, to

be predicted and characterized in real-time via broad molecular profiling of tissue samples.

Unlike other metabolomic approaches that require tedious sample preparation and analysis times,

this technology could be further developed as an on-line system in the processing environment to

enable meaningful sorting of beef products into categories reflecting tangible differences in

eating characteristics.

Current beef quality grading standards are applied from the visual assessment of two

carcass traits: marbling score and carcass maturity. Research has shown that these grading

standards generally separate carcasses based on predicted eating experiences (Smith et al., 1987; Platter et al., 2003; Emerson et al., 2013). In 2016, only 1.8% of graded fed beef carcasses had an overall USDA maturity score of C or greater (Boykin et al., 2017). This indicates that carcass maturity plays a very minimal role in determining quality grades in today's fed beef industry and that quality grade is almost entirely determined by marbling score. It is the general consensus that as marbling score increases, the probability of a positive eating experience also increases (Emerson et al., 2013). Although marbling is a major component of the grading system, it has shown to account for as little as 5% of variation in eating quality (Platter et al., 2003), clearly leaving significant sources of variation unaccounted for during the grading process. Molecular and biochemical components of beef muscle are known to influence beef eating quality and may explain variation not accounted for by marbling score alone (Mottram, 1998), but cannot be visually assessed by a human grader or grading camera. Therefore, the objective of this study was to evaluate the ability of rapid evaporative ionization mass spectrometry as a novel method to predict various components of beef quality including: carcass type, sensory attributes, and objective tenderness measurements.

Materials and Methods

Institutional Animal Care and Use Committee approval was not required for this study as samples were obtained from federally inspected harvest facilities.

Sample Collection

Beef strip loin sections were collected to represent 7 carcass types [Select (n = 42), Low Choice (n = 42), Top Choice (n = 41), Prime (n = 42), dark cutter (n = 41), grass-fed (n = 42), and Wagyu (n = 42)] in order to provide significant variation in beef flavor attributes, tenderness, fat percentages, and animal background. Product specifications for each carcass type were

verified by Colorado State University (CSU) personnel using official USDA grades and personal communication with individual suppliers to verify origin. From each half of the carcass, a 5 cm long strip loin section was fabricated from a point starting at the 13th rib. Each strip loin section was fabricated into one 2.54 cm steak. The steak from the left strip loin section was assigned to trained sensory analysis, with the remaining portion reserved for rapid evaporative ionization mass spectrometry (REIMS). The steak from the right strip loin section was assigned to shear force analysis. Steaks were individually vacuum packaged, aged (34°C) for 14 d postmortem, and frozen (-20°C) until analysis. Selection criteria from each carcass type was as follows:

Select

Select sections were chosen from A maturity carcasses with Sl^{00}-Sl^{99} marbling scores and an overall USDA Select quality grade. Only carcasses presenting typical beef-type characteristics were included to avoid dairy-type and *bos indicus* influence. Additionally, carcasses were visually free of dark cutting characteristics.

Low Choice

Low Choice sections were chosen from A maturity carcasses with Sm^{00}-Sm^{99} marbling scores and an overall USDA Low Choice quality grade. Only carcasses presenting typical beef-type characteristics were included to avoid dairy-type and *bos indicus* influence. Additionally, carcasses were visually free of dark cutting characteristics.

Top Choice

Top Choice sections were chosen from A maturity carcasses with Mt^{00}-Md^{99} marbling scores and an overall USDA quality grade of either Average or High Choice. Only carcasses presenting typical beef-type characteristics were included to avoid dairy-type and *bos indicus* influence. Additionally, carcasses were visually free of dark cutting characteristics.

Prime

Prime sections were chosen from A maturity carcasses with $SlAb^{00}$-Ab^{99} marbling scores and an overall USDA Prime quality. Only carcasses presenting typical beef-type characteristics were included to avoid dairy-type and *bos indicus* influence. Additionally, carcasses were free of dark cutting characteristics.

Dark Cutter

Dark cutter sections were chosen from dark cutting A maturity carcasses with Sm^{00}-Md^{99} marbling scores. Only carcasses presenting typical beef-type characteristics were included to avoid dairy-type and *bos indicus* influence. Additionally, carcasses produced dark cutting characteristics.

Grass-fed

Grass-fed sections were chosen from lots of beef-type cattle known to have been fed a grass diet for the entirety of their lives. Additionally, carcasses were A maturity with Sm^{00}-Md^{99} marbling scores and an overall USDA Choice quality grade. Additionally, carcasses were visually free of dark cutting characteristics.

Wagyu

Wagyu sections were selected from crossbred Wagyu cattle (50% Wagyu, 50% Angus). Carcasses were A maturity with $SlAb^{00}$-Ab^{99} marbling scores and were visually free of dark cutting characteristics.

Trained Sensory Analysis

Before cooking, frozen sensory steaks were thawed at 0-4°C for 16-24 h. Steaks were cooked in a commercial steam convection oven (Model SCC WE 61 E; Rational, Landsberg am Lech, Germany) set at 204°C and 0% humidity to a peak internal temperature of 71°C. Steak

temperature was monitored during the cooking process in the geometric center of a single steak using the oven core temperature probe (Model SCC WE 61 E; Rational, Landsberg am Lech, Germany). After cooking, the peak temperature of each steak was measured in the geometric center of each steak using a probe thermometer (SPLASH-PROOF SUPER-FAST® THERMAPEN®, ThermoWorks, Lindon, UT). Steaks were trimmed of all external fat and connective tissue, sized into 1 cm^2, and served to trained panelists. All panelists were trained to evaluate tenderness, juiciness, beef flavor ID, browned, roasted, metallic, fat-like, buttery, umami, sour, bitter, burnt, livery, green/hay-like, and rancid flavor attributes on a 100 mm unstructured line scale verbally anchored at both end (0 = extremely tough, extremely juicy, not present; 100 = extremely tender, extremely juicy, extremely intense) adapted from the beef flavor lexicon described by Adhikari et al. (2011). Attributes and references are presented in Table 3.1.

Two sensory sessions were conducted a day. For each session, six panelists evaluated 10-11 individual samples over 28 sensory sessions. Samples were assigned randomly to panels so that three complete replications were analyzed per sensory day, with 11 samples being served on odd panel sessions and 10 samples being served on even panel sessions. Before sample evaluation, a warm-up sample was served at the beginning of each session in order to keep panelists calibrated. Warm-up samples were USDA Low Choice strip steaks cooked using the same parameters as experimental steaks. A consensus among warm-up steak attributes was agreed upon by panelists before moving forward with remaining samples.

Shear Force

Both Warner-Bratzler (WBS) and slice shear force (SSF) measurements were obtained from every steak using procedures described by Lorenzen et al. (2010). Within 5 min of

recording peak internal temperature, a 1 cm × 5 cm slice was removed from the steak parallel to the muscle fibers from the lateral end and sheared perpendicular to the muscle fibers, using a universal testing machine (Instron Corp., Canton, MA) equipped with a flat, blunt-end blade (crosshead speed: 500 mm/min, load capacity: 100 kg), resulting in a single SSF measurement for each steak. The remaining portion of each steak was allowed to equilibrate to room temperature (22°C) and at least 4 cores (1.2 cm in diameter) were removed from each steak parallel to the muscle fibers. Each core was sheared perpendicular to the muscle fibers using a universal testing machine (Instron Corp., Canton, MA) fitted with a Warner-Bratzler shear force head (crosshead speed: 200 mm/min, load cell capacity: 100 kg). Peak shear force of each core was recorded, and the resulting values were averaged to obtain a single WBSF measurement for each steak.

Rapid Evaporative Ionization Mass Spectrometry (REIMS)

Metabolomic fingerprint profiling of strip loin sections was performed using rapid evaporative ionization mass spectrometry (REIMS). Before analysis, REIMS samples were thawed at 0-4°C for 16-24 h. Sample were analyzed using a Synapt G2 Si Q-ToF, fitted with a REIMS ionization source coupled with an iKnife sampling device (Waters Corporation, Milford, MA). Five burns per sample were collected in negative ion mode, with each burn lasting roughly 1 sec. The burns from each sample were collected from a 2.54 × 2.54 cm square from the center of the steak (Figure 3.1). The relative abundance values from the 5 burns were averaged to create a single value for each sample. Data were collected in the mass range from100-1,000 m/z. Data were preprocessed to include lock mass correction (leucine enkephalin), background subtraction, and normalized to the total ion current. Additionally, individual peaks were binned in intervals of 0.5 m/z starting with 100.25 and ending with 999.75 for a total of 1,800 variables. Data binning

is a technique used to account for minor errors during data collection. Leucine enkephalin has a molecular weight of 555.632 g/mol and its detection would interfere with neighboring components. Therefore, mass bins in the range of 550-600 were excluded from the data matrix for a total of 1,700 mass bins used for analysis.

Statistical Methods

Trained Sensory Panel Ratings and Shear Force Values

Although the objective of the study was not to characterize sensory attributes or shear force values specific to the carcass types, mean sensory ratings and shear force values are provided for reference (Tables 3.2 and 3.3). Sensory data were fit to a linear mixed model to evaluate differences among carcass types using the lmer function of the lme4 package in R (R Core Team, 2018). Before analysis, individual panelist ratings were averaged per sample so that each sample was analyzed using a single value for each attribute. The model was fit using carcass type as the main effect, sensory session and sample feed order as random effects, and peak cooking temperature as a covariate. Shear force data was also analyzed using a linear mixed model with carcass type serving as the main effect and peak cooking temperature included as a covariate.

Composite Sensory Scores

In order to evaluate the ability of REIMS to predict samples with similar eating characteristics, principle component analysis (PCA) was used to calculate overall sensory scores using the PCA function of the FactoMineR package in R (R Core Team, 2018). Principal component analysis is a statistical tool that can be used to detect clusters of related samples in a multivariate data set. This method can take highly collinear multivariate data and finds uncorrelated linear relationships in a multi-dimensional space by calculating new latent

variables, called principal components. By default, the first principal component explains the greatest amount of variation within the data set, followed by the second component, and so forth. Additionally, PCA has the capacity to detect predictor variables that drive differences in the values of the newly calculated principal components. For each principal component, a coefficient is assigned to each variable and that coefficient is applied to its respective variable for each sample. It is the summation of these coefficient-multiplied variables that become the new latent variables, or scores. From a sensory standpoint, this allows for the ability to identify individual samples with similar or dissimilar overall sensory characteristics. Using these concepts of PCA, scores from principal components were used to assign samples into categories based on overall sensory characteristics. Two separate PCAs were calculated: one including all measured sensory attributes (overall) and one including only flavor attributes. Sensory responses were center scaled prior to modeling. Hierarchical clustering analysis was performed on the factor scores of the 1st 2 components of each model to identify groups of observations with similar characteristics. Each scores plot was clustered into 3 groups and identified as positive, neutral, and negative.

Predictive Model Descriptions

Several predictive models were evaluated to assess capabilities of REIMS for numerous beef grading and classification scenarios. Modeling objectives included: identification of carcass type, sensory prediction, and shear force tenderness classification. Seven total prediction models were evaluated: 1) carcass type, 2) overall sensory 3 class, 3) overall sensory 2 class, 4) flavor 3 class, 5) flavor 2 class, 6) WBS tenderness, 7) and SSF tenderness.

The original carcass type model was fit using the 7 individual carcass type classifications as described above. Due to severe overlap in the misclassification of Select and Low Choice, as well as, Top Choice and Prime predictions, the decision was made to combine these groups of

carcass types to create 2 new classification categories: Select/Low Choice and Top Choice/Prime. Additionally, in the current study, these treatments produced similar sensory responses from trained panelists (Table 2). Therefore, a reduced model to predict carcass type included 5 total classification categories: Select/Low Choice, Top Choice/Prime, Dark Cutter, Grass-fed, and Wagyu.

Both overall sensory and flavor models were fit using the 3 classes as described above. Additionally, a second reduced model was built for each by combining neutral and negative responses and labeling the aggregating class negative, resulting in a model with a binary positive/negative response. Sensory classes were reduced in this manner to identify the ability of REIMS Samples were classified as either tough or tender based on SSF and WBS values. Tender classification was assigned to any sample receiving a shear force value ≥ 3.9 kg or 15.4 kg for WBS or SSF, respectively (ASTM F2925-11). Tough classification was assigned to any sample receiving shear force values greater than the aforementioned tender threshold.

Model Building

Data preprocessing and splitting into training and testing sets was similar for each model. Models were built and classifications predicted using PLS-LDA. Using this method, PLS reduced dimensionality and collinearity within the data set before classification with the LDA model by using individual scores values from a predetermined number of PLS components as input for LDA. Before fitting the model, variables with correlation coefficients ≥ 0.90 were identified. Of the highly collinear pairs, one was removed from the data set, reducing the number of mass bins to 1,332. Mass bins were then log transformed to address skewness of data distributions, mean centered, and unit variance scaled so that each variable had a mean of zero and an equal distribution. Eighty percent of the pre-processed data were then randomly selected

to train the models, with the remaining 20% set aside to test the prediction accuracy of the newly developed models. Splitting of the data was performed separately for each model so that each classification category could be evenly distributed between training and testing sets. Prediction models were fit using the pls.lda function from the plsgenomics package in R (R Core Team, 2018). The number of PLS components used as inputs for LDA was determined to maximize the predictability of the test data set. The advantage in discriminating power of coupling PLS with LDA can be observed in Figure 3.2. Table 3.4 presents the number of PLS components used and prediction accuracies for the final models.

Measure of Predictive Abilities

Several measures were calculated to evaluate predictive ability of each model using the predictions of the model built with the training set on the test set. Two measures were calculated to evaluate the overall predictive abilities of each model: overall prediction accuracy and balanced prediction accuracy. Overall prediction accuracy was calculated as the number of true positives divided by the total number of samples over all classes. However, if class imbalance exists in the data set, overall prediction accuracy can be biased towards dominant classes. For that reason, balanced prediction accuracy was also calculated for each model to provide a more realistic representation of a model's predictive ability. Balanced prediction accuracy was calculated as the average accuracy of each class. Sensitivity and precision were calculated on a class-by-class basis. Sensitivity is determined as the number of true positives divided by the number of true positives plus false negatives. It is essentially the accuracy of each individual class. Sensitivity does not, however, take false positives into consideration; therefore, precision was calculated for each class in conjunction with sensitivity. Precision is equal to the number of true positives divided by the number of true positives plus false positives.

Results and Discussion

Carcass Type Classification

Balanced prediction accuracy was 59.7% (Table 3.5) when all 7 carcass types were included as response variables in the prediction model. Wagyu carcasses were predicted with 100.0% sensitivity and 100.0% precision. Therefore, not only was REIMS output able to correctly identify every Wagyu carcass, but it also did not misclassify any carcasses as Wagyu. Grass-fed carcasses were also predicted with 100.0% sensitivity; however, they had a decreased rate of precision (80.0%). REIMS output correctly identified each Grass-fed carcass as such; however, it also produced 2 false positive Grass-fed carcasses (1 Low Choice and 1 Prime). Dark Cutters were predicted with 75.0% sensitivity, but showed increased precision (85.7%). These carcasses were selected to have a wide range in the severity in the visual appearance of dark cutting characteristics. Out of all carcass types, Dark Cutters provided the greatest variation in sensory ratings. Therefore, misclassification of this treatment group may have been related to the severity of the dark cutting condition.

Although misclassification was minimal for Wagyu, Grass-fed, and Dark Cutter carcass types, the carcass types associated with the general USDA quality grades (Select, Low Choice, Top Choice, and Prime) suffered from severe misclassification rates. Of the 22 carcasses misclassified, 17 of the misclassifications occurred within Select, Low Choice, Top Choice, and Prime carcasses. Among these treatments, the model showed the greatest predictive ability to correctly predict Prime carcasses. Prime carcasses were predicted with 50.0% sensitivity and 66.7% precision. Interestingly, the model showed propensity to classify unknown samples as Top Choice, with 12 of 55 carcasses being predicted as Top Choice. Furthermore, of these 12 samples, 9 were false positives. Both Select and Low Choice carcasses were predicted with

40

25.0% sensitivity and 33.3% precision. As shown by the projection of linear discriminant scores on the training set, the model easily produces clusters of Wagyu, Grass-fed, Dark Cutter, and Prime carcasses; however, severe overlap remains among Select, Low Choice, and Top Choice within the first 4 linear discriminants (Figure 3.2).

In addition to misclassification of Select, Low Choice, Top Choice, and Prime carcasses, trained sensory means showed only minor differences between Select and Low Choice carcasses, as well as, between Top Choice and Prime carcasses (Table 3.2). Although each group of carcasses would provide differences in merchandising value, only minimal differences in marbling score exist between adjacent USDA quality grades. Eating quality generally improves as marbling score increases (Emerson et al., 2013); however sensory similarities among neighboring quality grades can occur (Martinez et al., 2017). In the current study, Prime and Top Choice steaks were rated similarly ($P > 0.05$) for tenderness, juiciness, beef flavor ID, browned, buttery, metallic, umami, sour, and green/hay-like flavor notes. Low Choice and Select steaks were rated similarly ($P > 0.05$) for tenderness, browned, fat-like, buttery, umami, sour, and green/hay-like flavor notes. The remaining carcass types (Dark Cutter, Grass-fed, and Wagyu) provided unique sensory profiles to each other. For several attributes, Dark Cutter carcasses were similar ($P > 0.05$) to Select, including tenderness, beef flavor ID, browned, buttery, umami, livery, and green/hay-like. However, Dark Cutter carcasses had the greatest ($P < 0.01$) rancid flavor intensity of all 7 carcass types. Additionally, Dark Cutter carcasses had greater ($P < 0.01$) fat-like intensity and lower ($P < 0.05$) metallic and sour intensities compared to Select carcasses.

Therefore, a second prediction model was built to evaluate REIMS' ability to predict carcass type when stratified by similarities in sensory attributes. This second model combined Select and Low Choice, as well as, Top Choice and Prime before model fitting, for a total of 5

carcass type categories: 1) Select/Low Choice, 2) Top Choice/Prime, 3) Dark Cutter, 4) Grass-fed, and 5) Wagyu. Re-fitting the model to these classes increased balanced prediction accuracy to 83.8% (Table 3.6), with particular improvement in the prediction of Select/Low Choice and Top Choice/Prime carcasses. As shown in Figure 3.3, the first 4 linear discriminants of the reduced training model clearly separated each of the 5 classes. Select/Low Choice samples were predicted with 87.5% sensitivity and 73.6% precision. The 2 misclassified Select/Low Choice samples were predicted as Top Choice/Prime. However, the model predicted 5 false positive Select/Low Choice samples including 3 Dark Cutters, 1 Top Choice/Prime, and 1 Grass-fed. Only a single Top Choice/Prime sample was misclassified, which was predicted to be Select/Low Choice. Similar to the previous model, Wagyu carcasses were predicted with both 100.0% sensitivity and precision, clearly showing a unique metabolomic fingerprint as identified by REIMS. Grass-fed and Dark Cutter carcasses were predicted with 100.0% precision, indicating that no carcasses were misclassified as either carcass type. However, 1 Grass-fed carcass was misclassified as Select/Low Choice and 3 Dark Cutters were misclassified as Select/Low Choice; resulting in sensitivity of 87.5% and 62.5%, respectively.

Although both models used the same observations for Dark Cutter and Grass-fed carcass types, differences in prediction measurements differed between the 2 models. Depending on model parameters, a prediction model can be attracted to one category over the other, altering prediction accuracy of not only the entire model, but also within individual classes. Therefore, parameters could be altered to focus on a specific category or categories that may be of greater interest. In the case of a packer wanting to verify Grass-fed carcasses, model parameters could be altered to favor specificity and sensitivity in predicting Grass-fed beef, specifically. This would likely result in a low overall accuracy; however, it may better fit the objective of that specific

scenario. Although REIMS struggled to predict assigned carcass types, reducing the model by combining carcass types with similarities in mean sensory ratings increased balanced prediction accuracy by nearly 25 percentage points. Future work needs to be done to further evaluate this observation; however, it may suggest that REIMS has the ability to appropriately segregate carcasses based on sensory profiles as opposed to USDA quality grade. Although this proof of concept study does not provide prediction accuracies needed for industry application, it does present the ability of REIMS to identify metabolomic differences in meat tissues.

Current instruments approved to determine marbling scores do so by capturing an image of the entire ribeye exposed between the 12th and 13th ribs. REIMS, on the other hand, only captures a small a subsample (as little as a few cm) of the ribeye (Figure 3.1). Thus, REIMS is affected by the heterogeneity of marbling distribution, which would be expected to make an accurate marbling score prediction problematic. However, unlike human graders or grading cameras, REIMS provides a metabolomic fingerprint of the ribeye, allowing for discrimination based on factors including muscle lipid profiles, of both neutral and phospholipids. This type of online measurement could be beneficial in identifying specific eating quality and biological differences among carcasses. Although marbling score and USDA quality grade generally segregate carcasses into meaningful groups differing in eating quality attributes, variability still exists, especially with lower quality grades and marbling scores (Emerson et al., 2013). The current USDA quality grading system does not have the capacity to determine differences in characteristics including production background, breed-type, and other biochemical components of muscle tissue. At this point, the objective is not to replace the current grading system with REIMS, but rather use REIMS to augment the current grading system by providing the ability to

identify and predict beef quality parameters using carcass or muscle characteristics that cannot be visually assessed.

Wagyu, Grass-fed, and Dark Cutter carcasses were selected to have marbling scores within ranges to allow for the comparison with the USDA quality grades, so that variation stemming from total fat content would be minimized and discrimination could be based on other compositional components. Although Wagyu carcasses generally provide marbling scores that exceed those established by the USDA, those selected for the current study had marbling scores necessary for the USDA Prime quality grade ($SlAb^{00}-Ab^{99}$). Both Grass-fed and Dark Cutter carcasses were selected to have marbling scores necessary for the USDA Choice quality grade ($Sm^{00}-Md^{99}$). Previous studies have shown REIMS has the ability to correctly identify different beef breeds, as well as, different fish species (Balog et al., 2016; Black et al., 2017; respectively). Similar to past results, the current data provide high sensitivity in predicting Wagyu and Grass-fed beef. Although the exact genetic background of Prime carcasses is unknown, all carcasses were selected to be visually free of dairy-type or *bos indicus* influence. However, Wagyu carcasses were verified to have 50% Wagyu and 50% Angus genetics. Although the current study does not provide an adequate design to completely evaluate breed-type discrimination, future work to evaluate REIMS' ability to identify different extents of breed influence could provide beneficial. Grass-fed beef garners significant premiums at retail. For the month of June 2018, Grass-fed beef ribeye steaks received an average premium of $8.62 per pound over commodity beef at retail (USDA-AMS, 2018). Gathering such a significant premium, verification could be beneficial for both retailers and consumers. The ability to objectively validate Grass-fed beef by its metabolomic profile at the time of grading would provide packers with a verification process that would not rely on a paper certification process or the need to

closely monitor Grass-fed carcasses throughout the entire process of harvest, chilling, and grading.

Within both prediction models, the only misclassified Dark Cutter carcasses were predicted to either be Select or Select/Low Choice. According to the USDA grading standards, a dark cutting carcass can be discounted an entire quality grade depending on the severity of dark cutting characteristics (USDA, 2017). Dark Cutter carcasses were selected based on the visual appearance of dark cutting characteristics and marbling score, without regard to the applied USDA quality grade. Therefore, carcasses misclassified as Select could have had a final USDA quality grade as such. Because of the dark lean color, dark cutting beef is easily assessed by human and instrument graders. The ability for REIMS to identify general dark cutting carcasses is not of much immediate interest. However, the segregation of Dark Cutter carcasses from Grass-fed, Low Choice, and Top Choice carcasses provides evidence that there are metabolomic characteristics of dark cutting carcasses that could prove to be important biomarkers to aide in further understanding of the causes of dark cutting beef. Because dark cutting is a significant quality defect and dark cutting carcasses receive severe discounts, research has endeavored to identify animals susceptible to dark cutting to provide proactive preventative measures. However, prediction of dark cutting carcasses from live animal phenotype leaves much to be understood (Mahmood et al., 2016).

Overall Sensory Classification

Figure 3.4 provides PCA factor scores and loadings for the first 2 components built from tenderness, juiciness, and flavor sensory ratings. Half (50.4%) of the variation in sensory attributes was captured in the first component, which was primarily driven by tenderness and moderately driven by juiciness, fat-like, beef flavor ID, buttery, and umami flavor notes. The

scores plot shows a general distribution by carcass type based on mean values; however, there was no obvious clustering of treatments based on sensory ratings. This shows each carcass type had within treatment variability of sensory characteristics. Principal component scores plots show similarities within a data set, regardless of treatment assignment (Bro and Smilde, 2014). If two observations are plotted near one another, it is an indication that they have similar values for their combinations of all response variables. In Figure 3.4, closely placed observations would have similar ratings for all sensory attributes, thus indicating samples with similar eating characteristics overall. Hierarchical clustering of principal component scores allowed for a systematic determination of samples with similar factor scores. Clustering divides observations into groups so that each member within a group is strongly related to each other, but weakly related to observations within other groups (Astel et al., 2007). Therefore, clustering of principal components was determined to be an appropriate method to classify samples into groups based on projected sensory characteristics, with complete disregard to carcass type. The assignment of observations into overall sensory groups based on clustering is presented in Figure 3.5. Although Dark Cutter samples generally fall on the left side of the plot and Wagyu samples on the right side, comparison of the two factor scores plots show each carcass type had observations fall within each overall sensory category.

The overall sensory model with 3 classes (positive, neutral, and negative) produced a balanced prediction accuracy of 56.1% (Table 3.7). Sensitivity and precision decreased as samples were predicted from positive to negative. Although prediction accuracy was low, almost the entirety of misclassification occurred between neighboring classes (i.e., positive vs neutral and negative vs neutral). No positive samples were misclassified as negative; whereas, only 1 negative sample was misclassified as positive. In a scenario to identify top performing carcasses

and apply premiums, the current model would nearly guarantee the exclusion of low performing carcasses. In a similar situation in which low performing carcasses were being identified for exclusion, no top performing carcasses would be wrongly discounted. The ability for a packer or retailer to further guarantee a positive eating experience past the capabilities of USDA quality grade could provide advantages in added value, as well as increased trust in a brand and consumer loyalty.

Use of hierarchical clustering of principal components of trained sensory data to assign an overall eating classification has not been attempted before. Therefore, the classification system used would need to be validated by consumers before bold conclusions can be made. Consumer sensory panels were not conducted in the current study to validate this classification system, but Table 3.8 provides mean trained sensory scores based on the predicted overall sensory classification that was assigned during the prediction of the test data set from the training model. Although misclassification error was high during prediction, sensory attributes of samples into predicted classes showed logical separation. Samples classified as positive received greater ($P < 0.05$) ratings for tenderness, juiciness, beef flavor ID, fat-like, and umami and lower ratings ($P < 0.05$) for sour and green/hay-like attributes than samples classified as negative. Neutral samples were rated similarly ($P > 0.05$) to positive samples for beef flavor ID, sour, and green/hay-like, but similarly ($P > 0.05$) to negative samples for tenderness, juiciness, fat-like, buttery, and umami.

Plotting the linear discriminant scores of the training model show general separation of overall sensory classes, but noticeable overlap still exists (Figure 3.6). Although visualization of the training set showed decent separation, test error will be equal to the training error at best, but will usually be greater (Ghatak, 2017). Issues with overlapping observations plagues almost all

47

applied predictive models (Kuhn and Johnson, 2016). Even when mean values differ between treatments, it is not uncommon for individual observations to have similar characteristics with one or more observations from another treatment. Typically, it is within these regions of overlap were misclassification occurs (Xiong et al., 2010). In this scenario, predicted class probabilities are usually calculated to be nearly equal, but many models default to assigning an observation to the class with the greatest probability, even if there is only a miniscule difference. For example, in a binary response prediction problem, the model usually defaults to whichever class has a probability greater than 0.5. Depending on the objectives of the modeler, the minimum probability requirement can be increased to add further certainty that an observation truly belongs to an assigned class. Although this would exclude some true positives, it would further guarantee true negative samples were identified as such. Using the example of marketing beef with a guaranteed positive experience, increasing the minimum probability for class assignment would further guarantee a customer received beef with a guaranteed positive eating experience. Increasing the minimum probability for classification would inevitably restrict some positive beef from receiving premiums due to normal prediction error. But, it would further guarantee that a customer paying a premium for a positive eating experience would not be disappointed with a low performing product, thus earning customer trust, satisfaction, and repeated business.

Due to class overlap, a second overall sensory model was built by combining neutral and negative samples, creating a binary response of positive and negative. The reduced model provided a balanced prediction accuracy of 75.43% (Table 3.9). Negative samples were correctly classified with 89.7% sensitivity, whereas, positive samples were only classified with 61.1% sensitivity. If the current binary model were used in application, 26.6% of samples predicted as positive would actually be negative. This error rate would likely be too large for utilization into a

48

branded program or quality control setting. Although overall prediction accuracy was lower with the 3-class model, only 5.2% of samples predicted as positive were actually negative. This level of error would prove to be much more appropriate in segregating beef carcasses into eating quality groups, irrespective of marbling score. Although carcasses with modest degrees of marbling and greater have less than a 10% chance of providing a negative eating experience, acceptability becomes increasingly more variable as marbling score decreases (Smith et al., 2008). Therefore, a method to identify high and low performers, especially within lower quality grades, could add value and consistency in the way that beef is marketed. The current model includes beef representing a wide range of marbling scores, but future work would benefit from focusing purely on predicting variation in consumer acceptability of lower quality beef.

Flavor Prediction

The same approach to assign samples into overall sensory classes was used to assign samples into flavor classes using hierarchical clustering of PCA factor scores. However, for flavor class assignment, tenderness and juiciness ratings were excluded from the PCA model in order to more precisely evaluate the ability of REIMS to predict flavor differences. The flavor PCA explained 43.4% of variation in the first component (Figure 3.7). Slightly greater trends in clustering of samples by treatment were observed, particularly with the general clustering of Wagyu carcasses in the bottom right quadrant, which is strongly influenced by fat-like, buttery, and umami flavor attributes. Beef flavor ID intensity provided the largest influence on factor scores in component 1, with browned, roasted, fat-like, buttery, and umami having moderate influence on the projection of factor scores in the same direction of component 1. Sour and metallic had the strongest negative contribution to component 1. Within component 2, roasted and fat-like showed a strong negative relationship to one another. Although it is likely that there

is a moderate association between tenderness and flavor attributes, there is a shift in the contribution of the individual flavor attributes between the overall sensory and flavor PCAs. The determination of flavor classes (positive, neutral, and negative) by cluster analysis is presented in Figure 3.8.

Similar to the overall sensory model, balanced prediction accuracy of the flavor model was 55.6% (Table 3.10). Projection of the linear discriminants generally separated classes, but, again, significant regions of overlap remained among all classes (Figure 3.9). Of the 3 classes, neutral samples were classified with the greatest sensitivity (66.7%). Positive prediction sensitivity was only 55.0%; however, no negative samples were predicted as positive. Similar to the overall sensory model, overall balanced prediction accuracy was low, but in a scenario to select top performing carcasses, the 3-class flavor model would successfully exclude all negative samples from a positive classification. In a similar manner as stated above, a model of this nature could appropriately assign carcasses to a group free of low performing flavor samples. This model showed greater difficulty in identifying negative samples. Negative sensitivity was only 44.4%, with a precision of 40.0%. When sensory scores were analyzed between samples based on their predicted assignment into flavor classes, those predicted as positive had greater ($P < 0.05$) browned, fat-like, buttery, and umami intensities and lower ($P < 0.05$) metallic intensities than those samples predicted as negative (Table 3.11).

Again, a second flavor model was built combining neutral and negative samples, creating a binary response of positive and negative. The reduced model provided a balanced prediction accuracy of 70.3% (Table 3.12). Negative samples were predicted with 79.5% sensitivity, whereas, positive samples were predicted with 61.1% sensitivity. Of the samples predicted as positive, over 40.0% were actually negative. With this level error in predicted positive samples,

this model would not provide the industry with enough confidence to predict flavor in this manner. In the current study, REIMS presents a slight advantage in predicting an overall sensory classification when tenderness, juiciness, and flavor attributes are included in the model. This indicates that there is some level of independence between factors affecting tenderness and juiciness, and those affecting flavor alone. Partial justification for the observed difference in accuracy may be explained by the eigenvalues of the principal components. A greater amount of variation was explained by the first 2 components of the overall model in comparison to the flavor model (67.0% vs 60.6%, respectively) that was used in the cluster analysis. Regardless, methods used to assign classes for both models left significant portions of variation unaccounted for. Evaluation of other methods to assign a composite trained sensory score could prove beneficial for prediction accuracy.

Tenderness Classification

Figures 3.10 and 3.11 show the distribution of SSF and WBS designated into tender and tough classes, respectively. REIMS metabolomic profiles were able to predict SSF and WBS tenderness categories with comparable overall accuracies (75.4% and 70.2%, respectively; Tables 3.13 and 3.14). Slice shear force is a more accurate and simpler method of predicting tenderness (Wheeler et al., 2004); thus it would be expected for SSF to be more predictable. Tender samples were predicted with a precision of 78.3%, showing potential for REIMS to effectively segregate carcasses into tender groups. REIMS predicted tough samples with 83.3% sensitivity, but with a decreased precision (73.5%). The WBS model had lower precision in predicting both tender and tough samples compared to the SSF model (72.4% and 67.8%, respectively), making the WBS less effective in predicting beef tenderness. Furthermore, with a smaller relationship to sensory tenderness, the WBS model would be expected to provide further

error in predicting consumer tenderness (Shackelford et al., 1999). Tenderness is a highly valued beef attribute and studies indicate consumers' ability to detect tenderness differences and a willingness to pay premiums for tender beef (Boleman et al., 1997; Miller et al., 2001; Platter et al., 2013). The USDA has standards in place for guaranteed tender programs based on SSF values (ASTM F2925-11), but it is perceived as too time consuming and destructive for industry application (Wheeler et al., 2002). Several instrument methods to predict tenderness have been evaluated that are less destructive and could be implemented at line speeds, but have yet to be developed for industry use (Park et al., 1998; Belk et al., 2001; Vote et al., 2003; Konda et al., 2008).

Of these methods, near-infrared reflectance (NIR) has produced the greatest results. Park et al. (1998) segregated samples into 2 WBS categories: < 6 kg and > 6 kg. The authors reported a similar overall accuracy as reported in the current study (79% vs 75%, respectively). However, in contrast to the current study, Park et al. (1998) had greater sensitivity in classifying tender samples and lesser sensitivity in classifying tough samples. Additionally, the cutoff WBS shear force value chosen is well above the ASTM's determination of a tough sample (4.4 kg). Therefore, it is difficult to determine the ability of their model to appropriately segregate tough and tender samples. Price et al. (2008) also evaluated the ability of NIR in predicting beef tenderness using cutoff values similar to those used in the current study for SSF (tender = > 16 kg, intermediate = 16 to 25 kg, and tough = < 25 kg; Price et al., 2008). The authors reported NIR accurately predicted 92.9% of tough samples; however, the paper makes no reference to the prediction accuracies of the other 2 tenderness categories.

More recently, hyperspectral imaging has gained attention as an instrument to predict beef tenderness with high accuracy. Calculated as the percent of truly tender samples out of all

samples predicted tender, Naganathan et al. (2016) reported tenderness certification accuracy of 87.6%. However, this accuracy metric is terribly misleading due to tenderness class imbalance. The model reported misclassified 41.0% of tender samples and 40% of tough samples, showing a severe lack of both sensitivity and specificity of the instrument in identifying tough samples (Konda Naganathan et al., 2016). A similar study published with collaboration from several authors of the Naganathan et al. (2016) paper reported an overall tenderness prediction accuracy of 89.8% using hyperspectral imaging (Nubiato et al., 2018). In this study, no reference of segregation of samples into a training and test was mentioned and it is assumed that accuracy measurements were calculated by cross-validation, which is generally an overestimation of model robustness. Additionally, no reference is made to the distribution of tough and tender samples. Therefore, caution needs to be taken when comparing the effectiveness of REIMS to hyperspectral imaging.

Conclusion

Eating satisfaction is critical for consumers to repeatedly purchase beef and increase consumer demand. Various avenues (i.e., the USDA grading system and branded beef programs) have been extensively developed to segregate beef that will provide consistent eating characteristics, so consumers feel confident in the beef they purchase. Although these programs have become very successful and generally meet this objective, it is not uncommon for consumers to have negative eating experiences, even when purchasing quality grades or brands they trust. Today, carcasses are graded and assigned into branded programs mainly from visual assessments of carcass characteristics and other indicators of animal age, background, or breed. However, there are numerous molecular components that influence beef eating quality that cannot be visually assessed or measured by a camera. Technologies like gas or liquid

chromatography, nuclear magnetic resonance, histology, and Raman spectroscopy allow for the ability to measure some of these molecular attributes; however, they require laborious sample preparation, long analysis times, or sometimes both. REIMS is unique to these technologies in its ability to provide a detailed and high-resolution metabolic profile in a matter of seconds without the need for sample preparation, allowing for its adaption into an on-line application.

The current project provides an initial evaluation for the efficacy of REIMS technology to determine various beef carcass characteristics. This relatively new mass spectrometry method is the first of its kind to exhibit the ability to completely process a metabolomic fingerprint from sample collection to classification prediction within seconds, making it a viable consideration for on-line application in beef packing facilities. The current study shows REIMS can predict carcass type, shear force values, and general sensory characteristics with moderately high accuracy. Previous studies evaluating the predictive capacity of REIMS have reported nearly perfect accuracy in their respective classification problems (Balog et al., 2010; St John et al., 2017; Verplanken et al., 2017). However, these studies evaluated responses with more defined differences than many of those used in the current study. Nevertheless, the current study showed REIMS to have the capability to identify Grass-fed and Wagyu carcasses with 100% accuracy. Classification accuracy of other beef carcass types was improved when treatments were stratified into groups with similar mean trained sensory attributes, suggesting an ability to more appropriately determine sensory differences, regardless of carcass type. Furthermore, REIMS identified tough versus tender samples based on SSF values with accuracies comparable, if not more favorable, to other beef quality prediction instrument candidates.

The current study presented issues with overlapping samples and cases of class imbalance, issues that plague classification efforts across various industries and applications.

Future evaluation of appropriate statistical and prediction approaches to appropriately address these issues could improve the predictive ability of REIMS. Furthermore, annotation and identification of compounds of interest could provide insight into the metabolic components related to various beef quality attributes. Regardless, the current study shows REIMS technology has the ability to provide a molecular fingerprint of muscle that is useful in predicting beef quality characteristics in a way that can be developed for on-line application and potentially provide a more meaningful approach to sort carcasses and apply premiums to various products. The information gained from this initial evaluation will allow for more focused studies in the future.

Table 3. 1. Definition and reference standards for beef descriptive sensory attributes and their intensities rated on a continuous line scale from 0 to 100 adapted from Adhikari et al. (2011).

Attribute	Definition	Reference
Tenderness	The overall tenderness of the sample.	Brisket steak to 160°F = Strip loin steak to 160°F = 60 Tenderloin to 160°F = 95
Juiciness	The amount of perceived juice that is released from the product during mastication.	Carrot = 55 Strip steak cooked to 175°F = 60 Strip Steak cooked to 135 = 75 Watermelon = 95
Beef Flavor ID	Amount of beef flavor identity in the sample.	Swanson's Beef Broth = 35 Beef Brisket = 80
Browned	Aromatic associated with the outside of grilled or broiled meat; seared but not blackened or burnt.	Beef Suet (broiled) = 55
Roasted	Aromatic associated with roasted meat.	80% Lean Ground Chuck = 65
Fat-Like	The aromatics associated with cooked animal fat.	Hillshire Farms lit'l Beef Smokies = 45 Beef suet (broiled) = 80
Buttery	Sweet, dairy-like aromatic associated with natural butter.	Land O'Lakes Unsalted Butter = 45
Metallic	The impression of slightly oxidized metal, such as iron, copper, and silver spoons.	0.10% Potassium Chloride Solution = 10 Dole Canned Pineapple Juice = 40
Umami	Flat, salty, somewhat brothy. The taste of glutamate, salts of amino acids and other molecules called nucleotides.	0.035% Accent Flavor Enhancer Solution = 50
Sour	The fundamental taste factor associated with citric acid.	0.015% Citric Acid solution = 10 0.050% Citric Acid solutions = 25
Bitter	The fundamental taste factor associated with a caffeine solution.	0.02% Caffeine Solution = 25
Livery	The aromatics associated with cooked organ meat/liver.	Beef Liver = 50
Green/Hay-Like	Brown/green dusty aromatics associated with dry grasses, hay, dry parsley, and tea leaves.	Dry parsley = 40
Rancid	The aromatics commonly associated with oxidized fat and oils. These aromatics may include cardboard, painty, varnish, and fishy.	Wesson Vegetable Oil microwaved 3 min = 45 Wesson Vegetable Oil microwaved 5 min = 60

Table 3. 2. Trained sensory ratings[1] for beef strip steaks of varying quality treatments.

Treatment	Tenderness	Juiciness	Beef Flavor ID	Browned	Roasted	Fat-Like	Buttery	Metallic	Umami	Sour	Bitter	Livery	Green/Hay-Like	Rancid
Wagyu[2]	68.3^a	64.8^a	41.3^a	29.1^a	33.5^{ab}	14.2^a	7.2^a	5.1^d	6.6^a	3.1^c	1.7	0.3^d	0.8^d	1.9^c
Prime	64.3^{bc}	60.9^b	40.4^{ab}	28.0^{ab}	34.3^{ab}	10.1^b	4.2^b	6.7^c	4.9^b	4.2^b	1.7	0.4^d	1.1^d	2.0^{bc}
Top Choice	62.6^b	59.3^{bc}	40.9^a	26.7^{bc}	33.3^{ab}	9.0^c	3.7^{bc}	6.6^c	4.5^{bc}	4.6^b	1.7	0.6^{cd}	1.1^d	2.9^b
Low Choice	59.2^d	57.3^c	39.0^b	26.5^{cd}	34.6^a	7.1^{de}	2.4^{de}	7.7^b	3.8^{cd}	6.1^a	2.0	0.6^{bcd}	1.7^{bc}	2.5^{bc}
Select	57.2^d	55.0^d	37.1^c	25.2^{de}	33.1^b	6.2^e	1.7^e	8.6^a	3.3^d	6.9^a	1.6	1.0^{abc}	2.0^{ab}	2.4^{bc}
Grass-fed[3]	67.8^{ab}	58.9^{bc}	39.0^b	25.2^{de}	34.4^{ab}	8.0^{cd}	3.1^{de}	7.0^{bc}	4.7^b	4.7^b	1.6	1.2^a	1.3^{cd}	1.8^c
Dark Cutter[4]	57.9^d	58.0^c	35.6^c	24.7^e	31.6^c	7.5^d	2.5^{de}	6.3^c	3.8^{cd}	4.1^b	2.4	1.0^{ab}	2.3^a	4.4^a
P – Value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.06	< 0.01	< 0.01	< 0.01
SEM[5]	1.3	0.9	0.8	0.7	0.6	0.4	0.3	0.3	0.3	0.4	0.2	0.2	0.2	0.4

^{abcde}Least square means in the same row without a common superscript differ ($P < 0.05$) due to treatment.

[1]Attributes were scored using an unstructured line scale anchored at both ends: 0 = very tough, very dry, and not present; 100 = very tender, very juicy, and very intense

[2]Wagyu strip steaks were selected from carcasses with Slightly Abundant^{00} - Abundant^{99} marbling scores.

[3]Grass-fed strip steaks were selected from carcasses with Small^{00} - Modest^{99} marbling scores.

[4]Dark Cutter strip steaks were selected from carcasses with Small^{00} - Modest^{99} marbling scores.

[5]Standard error (largest) of the least squares means.

Table 3. 3. Slice shear force (SSF) and Warner-Bratzler shear force (WBS) values of beef strip steaks cooked to 71°C from various carcass types.

Treatment	SSF	WBS
Wagyu[1]	16.77[b]	3.60[d]
Prime	16.08[bc]	3.97[cd]
Top Choice	17.48[b]	4.10[bcd]
Low Choice	18.27[b]	4.54[b]
Select	17.23[b]	4.30[bc]
Grass-fed[2]	13.94[c]	3.03[e]
Dark Cutter[3]	22.29[a]	5.10[a]
P – Value	< 0.001	< 0.001
SEM[4]	0.89	0.20

[abcde]Least square means in the same row without a common superscript differ ($P < 0.05$) due to treatment.
[1]Wagyu strip steaks were selected from carcasses with Slightly Abundant[00] - Abundant[99] marbling scores.
[2]Grass-fed strip steaks were selected from carcasses with Small[00] - Modest[99] marbling scores.
[3]Dark Cutter strip steaks were selected from carcasses with Small[00] - Modest[99] marbling scores.
[4]Standard error (largest) of the least squares means.

Table 3. 4. Prediction model outline for various beef quality attributes produced from metabolomic profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry (REIMS).

Model[1]	Classification Categories	Number of PLS Components	Overall Prediction Accuracy, %	Balanced Prediction Accuracy, %
Carcass Type	Wagyu Prime Top Choice Low Choice Select Grass-fed Dark Cutter	18	60.0	59.7
Combined Carcass Type	Wagyu Prime/Top Choice Low Choice/Select Grass-fed Dark Cutter	13	87.5	83.8
Overall Sensory (3-Class)	Positive Neutral Negative	7	59.7	56.1
Overall Sensory (2-Class)	Positive Negative	5	80.7	75.43
Flavor (3-Class)	Positive Neutral Negative	7	59.7	55.6
Flavor (2-Class)	Positive Negative	4	73.8	70.3
Slice Shear Force	Tender Tough	7	75.4	75.0
Warner-Bratzler Shear Force	Tender Tough	9	70.2	70.2

[1]Each model was fit using partial least squares (PLS) as a dimension reduction technique coupled with linear discriminant analysis (LDA) for classification.
[2]Number of PLS components included as predictor variables for LDA

Table 3. 5. Misclassification matrix[1] of various beef carcass types as predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry.

				Predicted Class						
Reference Class	Select	Low Choice	Top Choice	Prime	Dark Cutter	Grass-fed	Wagyu	Total	Sensitivity	Precision
Select	2	3	2	1	0	0	0	8	25.0%	33.3%
Low Choice	0	2	5	0	0	1	0	8	25.0%	33.3%
Top Choice	1	1	3	1	1	0	0	7	42.8%	25.0%
Prime	1	0	2	4	0	1	0	8	50.0%	66.7%
Dark Cutter	2	0	0	0	6	0	0	8	75.0%	85.7%
Grass-fed	0	0	0	0	0	8	0	8	100.0%	80.0%
Wagyu	0	0	0	0	0	0	8	8	100.0%	100.0%
Total	6	6	12	6	7	10	8	55		

Overall Prediction Accuracy   60.0%
Balanced Prediction Accuracy   59.6%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.

Table 3. 6. Misclassification matrix[1] of various beef carcass types as predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry.

| Reference Class | Predicted Class | | | | | Total | Sensitivity | Precision |
	Select/Low Choice	Top Choice/Prime	Dark Cutter	Grass-fed	Wagyu			
Select/Low Choice	14	2	0	0	0	16	87.5%	73.6%
Top Choice/Prime	1	15	0	0	0	16	93.7%	88.2%
Dark Cutter	3	0	5	0	0	8	62.5%	100.0%
Grass-fed	1	0	0	7	0	8	87.5%	100.0%
Wagyu	0	0	0	0	8	8	100.0%	100.0%
Total	19	17	5	7	8	56		

Overall Prediction Accuracy	87.5%
Balanced Prediction Accuracy	83.7%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.

Table 3. 7. Misclassification matrix[1] of 3 overall sensory categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry (REIMS).

	Predicted Class					
Reference Class[3]	Positive	Neutral	Negative	Total	Sensitivity	Precision
Positive	13	5	0	18	72.2%	68.4%
Neutral	5	17	4	26	65.3%	56.6%
Negative	1	8	4	13	30.7%	50.0%
Total	19	30	8	57		

Overall Prediction Accuracy	59.6%
Balanced Prediction Accuracy	56.1%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Reference class assigned using hierarchical clustering of principal components of PCA data.
Positive = cluster 1, Neutral = cluster 2, Negative = cluster 3.

Table 3. 8 Trained sensory ratings[1] for beef strip steaks predicted into overall sensory classes2 using mass spectra collected with rapid evaporative ionization mass spectrometry (REIMS).

Attribute	Positive	Neutral	Negative	SEM[3]	P – Value
Tenderness	69.0[a]	60.7[b]	56.7[b]	2.5	< 0.05
Juiciness	62.2[a]	58.6[b]	56.0[b]	2.0	0.2
Beef Flavor ID	40.1[a]	40.1[a]	35.0[b]	1.5	0.01
Browned	27.9	27.1	24.7	1.3	0.14
Roasted	32.7	33.7	32.0	1.4	0.43
Fat-Like	11.6[a]	8.4[b]	6.9[b]	1.2	< 0.01
Buttery	5.3[a]	3.4[b]	1.9[b]	0.8	< 0.01
Metallic	5.8	7.1	6.7	0.8	0.18
Umami	5.6[a]	4.5[b]	3.6[b]	0.6	< 0.01
Sour	3.4[b]	5.2[b]	6.2[a]	0.9	0.23
Bitter	1.6	1.7	1.1	0.5	0.52
Livery	0.4	0.6	0.8	0.3	0.55
Green/Hay-Like	1.1[b]	1.1[b]	3.1[a]	0.4	< 0.01
Rancid	2.0	2.5	3.4	1.0	0.48

[ab]Least square means in the same row without a common superscript differ ($P < 0.05$) due to treatment.
[1]Attributes were scored using an unstructured line scale anchored at both ends: 0 = very tough, very dry, and not present; 100 = very tender, very juicy, and very intense
[2]Sensory classes were assigned using hierarchical clustering of principal component factor scores based on trained sensory ratings for tenderness, juiciness, and flavor attributes. Prediction models were fit using Partial Least Squares-Linear Discriminant Analysis.
[3]Standard error (largest) of the least squares means.

Table 3. 9 Misclassification matrix[1] of 2 overall sensory categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry (REIMS).

	Predicted Class				
Reference Class[3]	Positive	Negative	Total	Sensitivity	Precision
Positive	11	7	18	61.1%	73.3%
Negative	4	35	39	89.7%	83.3%
Total	15	42	57		

Overall Prediction Accuracy	80.7%
Balanced Prediction Accuracy	75.4%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Reference class assigned using hierarchical clustering of principal components of PCA data. Positive = cluster 1, Negative = clusters 2 and 3.

Table 3. 10 Misclassification matrix[1] of 3 flavor categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry (REIMS).

Reference Class[3]	Predicted Class			Total	Sensitivity	Precision
	Positive	Neutral	Negative			
Positive	10	6	2	18	55.5%	62.5%
Neutral	6	20	4	30	66.7%	64.4%
Negative	0	5	4	9	44.4%	40.0%
Total	16	31	10	57		

Overall Prediction Accuracy   59.6%
Balanced Prediction Accuracy   55.5%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Reference class assigned using hierarchical clustering of principal components of PCA data. Positive = cluster 1, Neutral = cluster 2, Negative = cluster 3.

Table 3. 11 Trained sensory ratings[1] for beef strip steaks predicted into flavor classes[2] using mass spectra collected with rapid evaporative ionization mass spectrometry (REIMS).

	Predicted Flavor Class				
Attribute	Positive	Neutral	Negative	SEM[3]	P - Value
Tenderness	64.8	61.3	59.4	2.9	0.29
Juiciness	59.5	57.3	57.1	1.6	0.51
Beef Flavor ID	41.4	38.9	36.3	1.6	0.05
Browned	29.0[a]	26.8[ab]	24.9[b]	1.2	0.03
Roasted	34.4	34.5	33.7	1.2	0.85
Fat-Like	11.6[a]	7.4[b]	6.3[b]	1.1	< 0.01
Buttery	5.6[a]	2.4[b]	2.2[b]	0.7	< 0.01
Metallic	6.5[b]	6.9[b]	8.7[a]	0.6	0.02
Umami	5.8[a]	4.1[b]	3.6[b]	0.5	< 0.01
Sour	4.3	5.5	6.0	0.9	0.25
Bitter	1.4	1.7	2.0	0.3	0.54
Livery	0.6	1.0	0.5	0.3	0.32
Green/Hay-Like	1.0	1.8	2.1	0.5	0.09
Rancid	2.4	2.6	4.2	0.9	0.24

[ab]Least square means in the same row without a common superscript differ ($P < 0.05$) due to treatment.
[1]Attributes were scored using an unstructured line scale anchored at both ends: 0 = very tough, very dry, and not present; 100 = very tender, very juicy, and very intense
[2]Sensory classes were assigned using hierarchical clustering of principal component factor scores based on trained sensory ratings for flavor attributes. Prediction models were fit using Partial Least Squares-Linear Discriminant Analysis.
[3]Standard error (largest) of the least squares means.

Table 3. 12 Misclassification matrix[1] of 2 flavor categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry (REIMS).

Reference Class[3]	Predicted Class		Total	Sensitivity	Precision
	Positive	Negative			
Positive	11	7	18	57.8%	61.1%
Negative	8	31	39	81.5%	79.4%
Total	19	38	57		

Overall Prediction Accuracy	73.6%
Balanced Prediction Accuracy	70.3%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Reference class assigned using hierarchical clustering of principal components of PCA data. Positive = cluster 1, Negative = clusters 2 and 3.

Table 3. 13 Misclassification matrix[1] of slice shear force (SSF) tenderness categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative

	Predicted Class				
Reference Class[3]	Tender	Tough	Total	Sensitivity	Precision
Tender	18	9	27	66.7%	78.3%
Tough	5	25	30	83.3%	73.5%
Total	23	34	57		

Overall Prediction Accuracy   75.4%
Balanced Prediction Accuracy   75.0%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Tender = SSF < 15.4; Tough = SSF ≥ 15.4
ionization mass spectrometry.

Table 3. 14 Misclassification matrix[1] of Warner-Bratzler shear force (WBS) tenderness categories predicted[2] by Partial Least Squares-Linear Discriminant Analysis using molecular profiles of beef strip steaks collected using rapid evaporative ionization mass spectrometry.

Reference Class[3]	Predicted Class		Total	Sensitivity	Precision
	Tender	Tough			
Tender	21	9	30	70.0%	72.4%
Tough	8	19	27	70.3%	67.8%
Total	29	28	57		

Overall Prediction Accuracy	70.1%
Balanced Prediction Accuracy	70.1%

[1]Number of samples falling into each respective classification category after prediction.
[2]Models were built using 80% of the original data and tested using the remaining 20%.
[3]Tender = WSF < 3.9; Tough = WSF ≥ 3.9

Figure 3. 1 Sampling location for rapid evaporative ionization mass spectrometry (REIMS) of beef strip steaks. Five burns were taken from each sample. Relative intensities of mass spectra from each burn were averaged to create a single data matrix per sample.
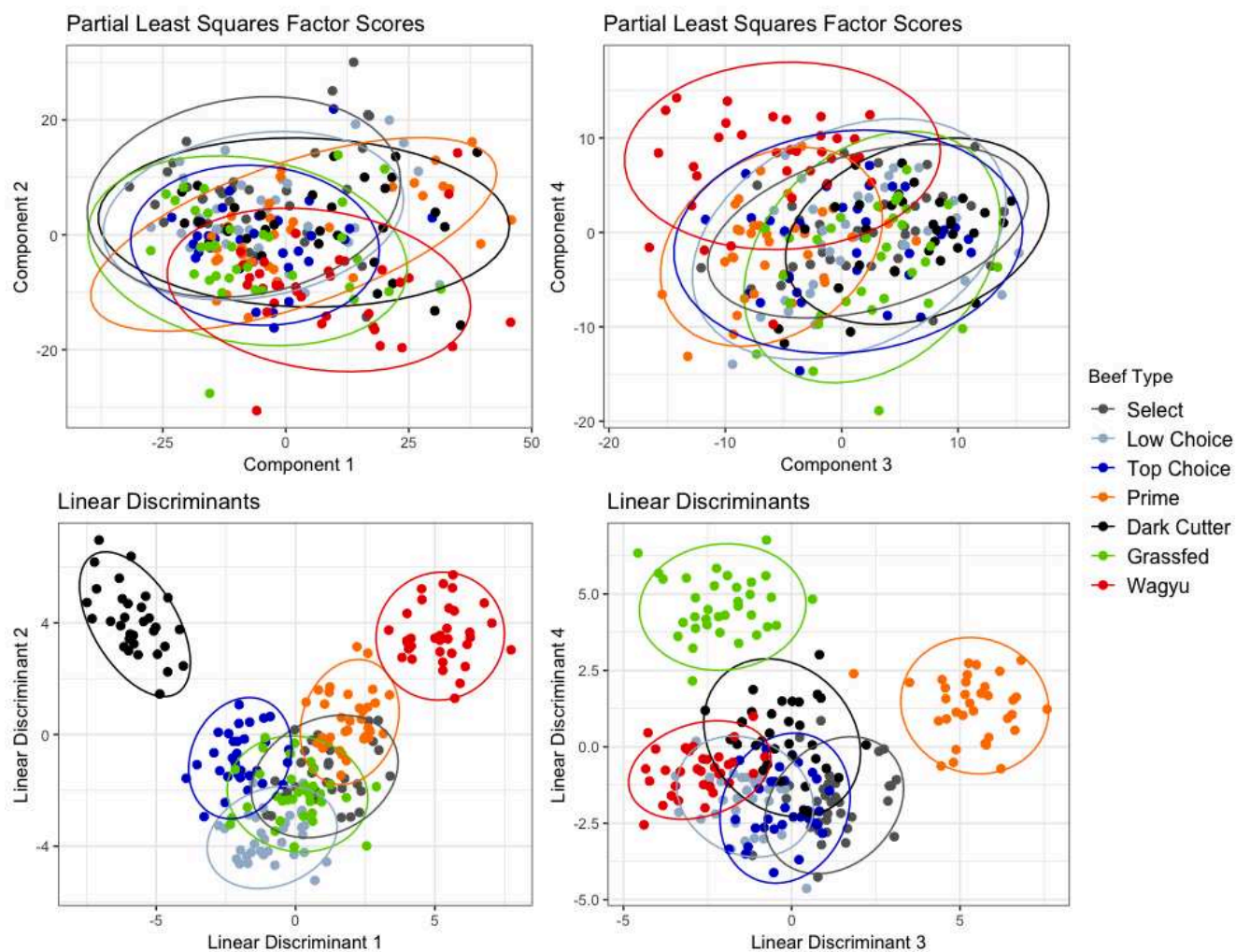
Figure 3. 2 Projection of partial least squares (PLS) scores (top) and linear discriminant (LDA) scores (bottom) of the training model built from rapid evaporative ionization mass spectrometry (REIMS) mass bins to predict various beef carcass types. Factor scores from PLS where used as inputs for LDA.

71

Figure 3. 3 Projection of partial least squares-linear discriminant scores of the training model built from rapid evaporative ionization mass spectrometry (REIMS) mass bins to predict various beef carcass types.
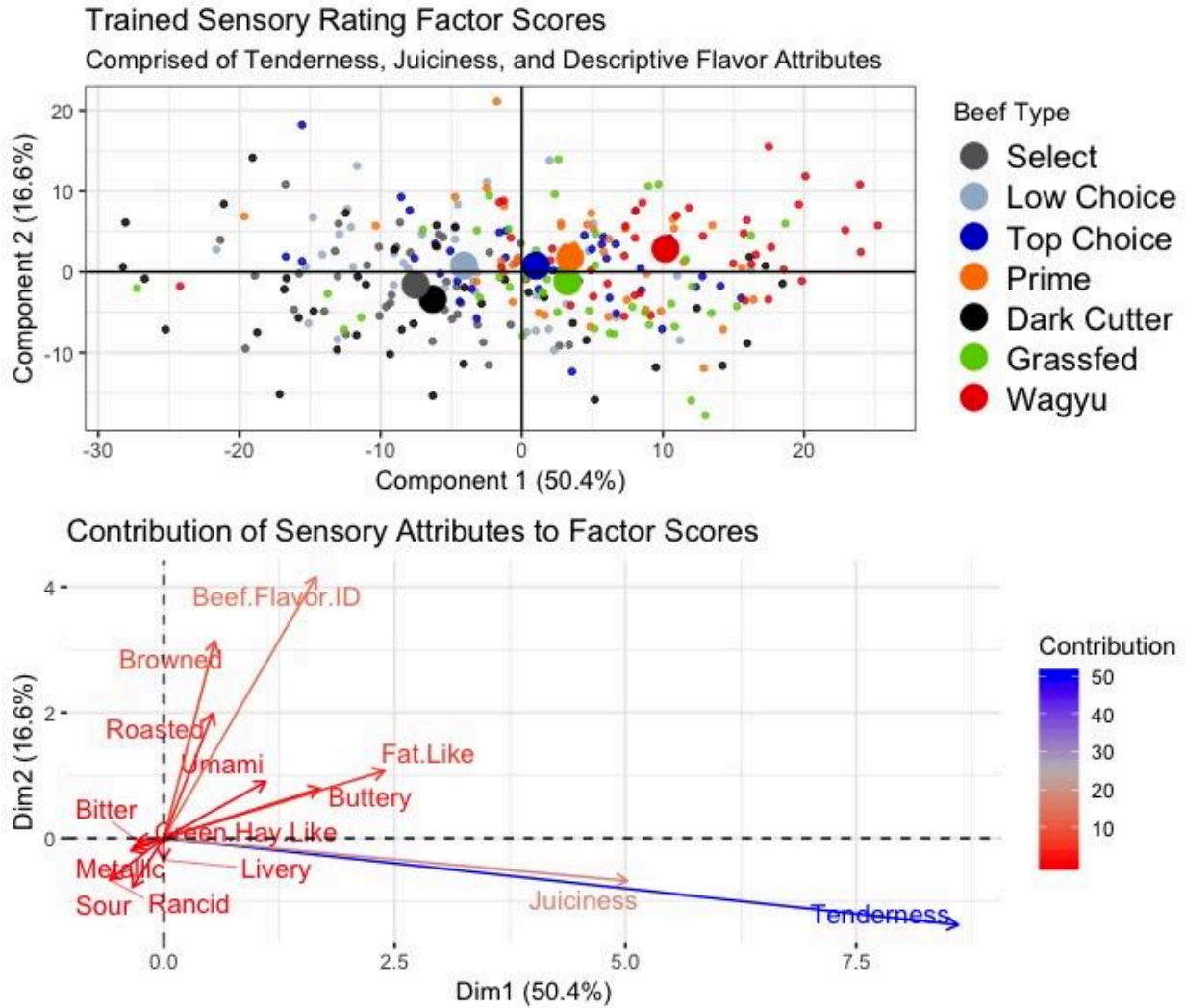
Figure 3. 4 (top) Projection of principal component scores of trained sensory ratings for tenderness, juiciness, and flavor attributes. Treatment centers are represented by large points. (bottom) Loadings plot showing the contribution of each sensory attribute to factor scores.
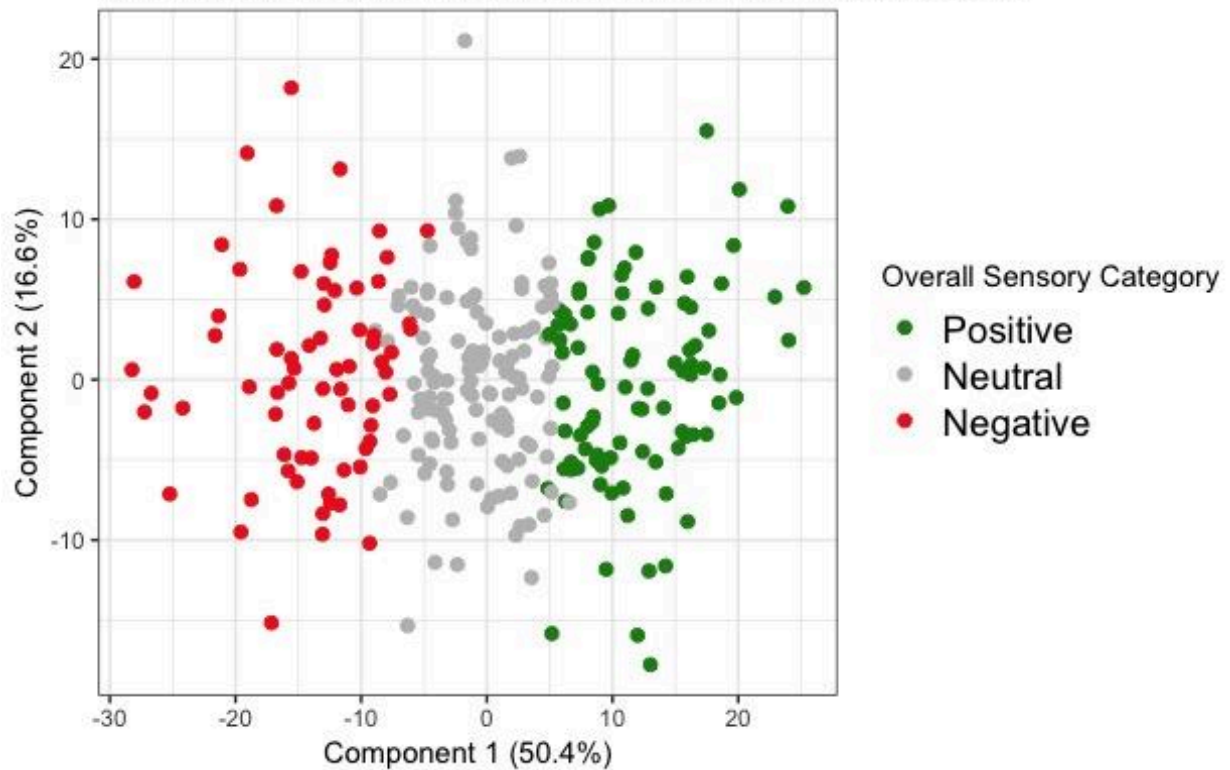
Figure 3. 5 Projection of principal component scores of trained sensory ratings for tenderness, juiciness, and flavor attributes. Scores are colored to represent overall sensory categories as determined by cluster analysis.
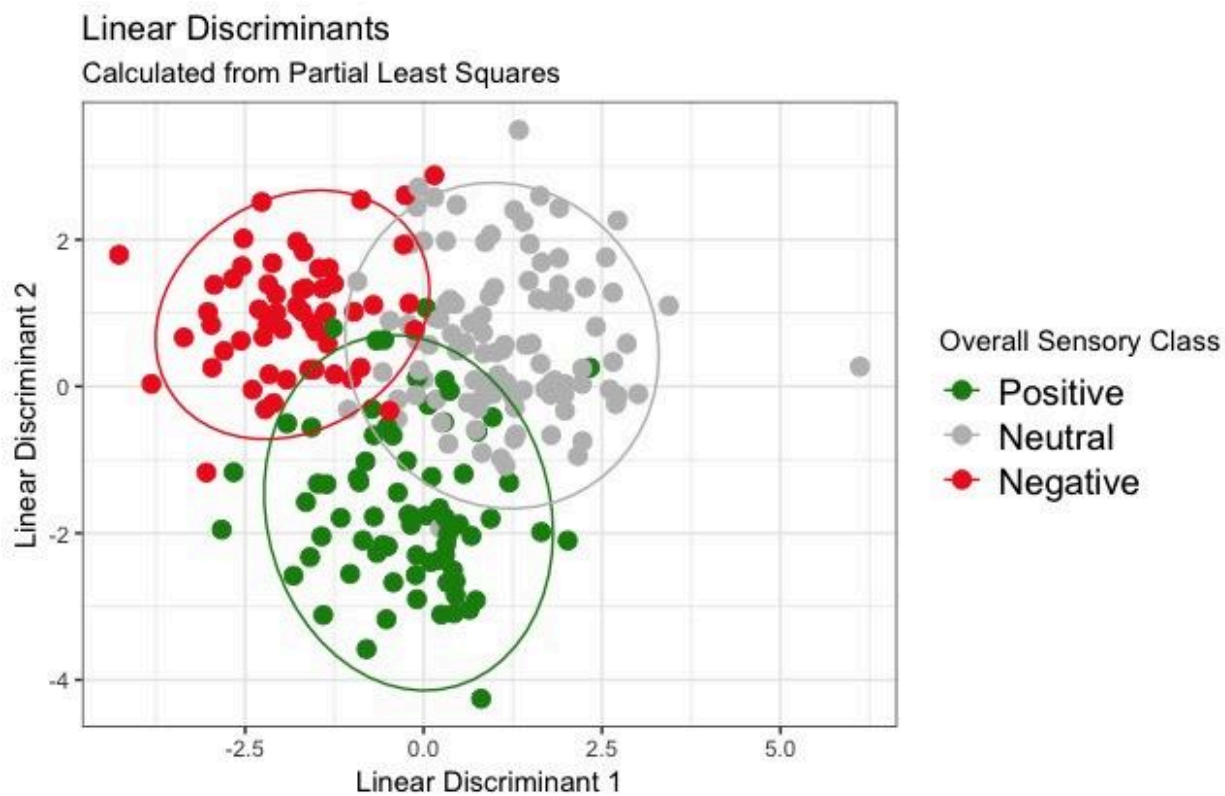
Figure 3. 6 Projection of partial least squares-linear discriminant scores of the training model built from rapid evaporative ionization mass spectrometry (REIMS) mass bins to predict overall sensory class.
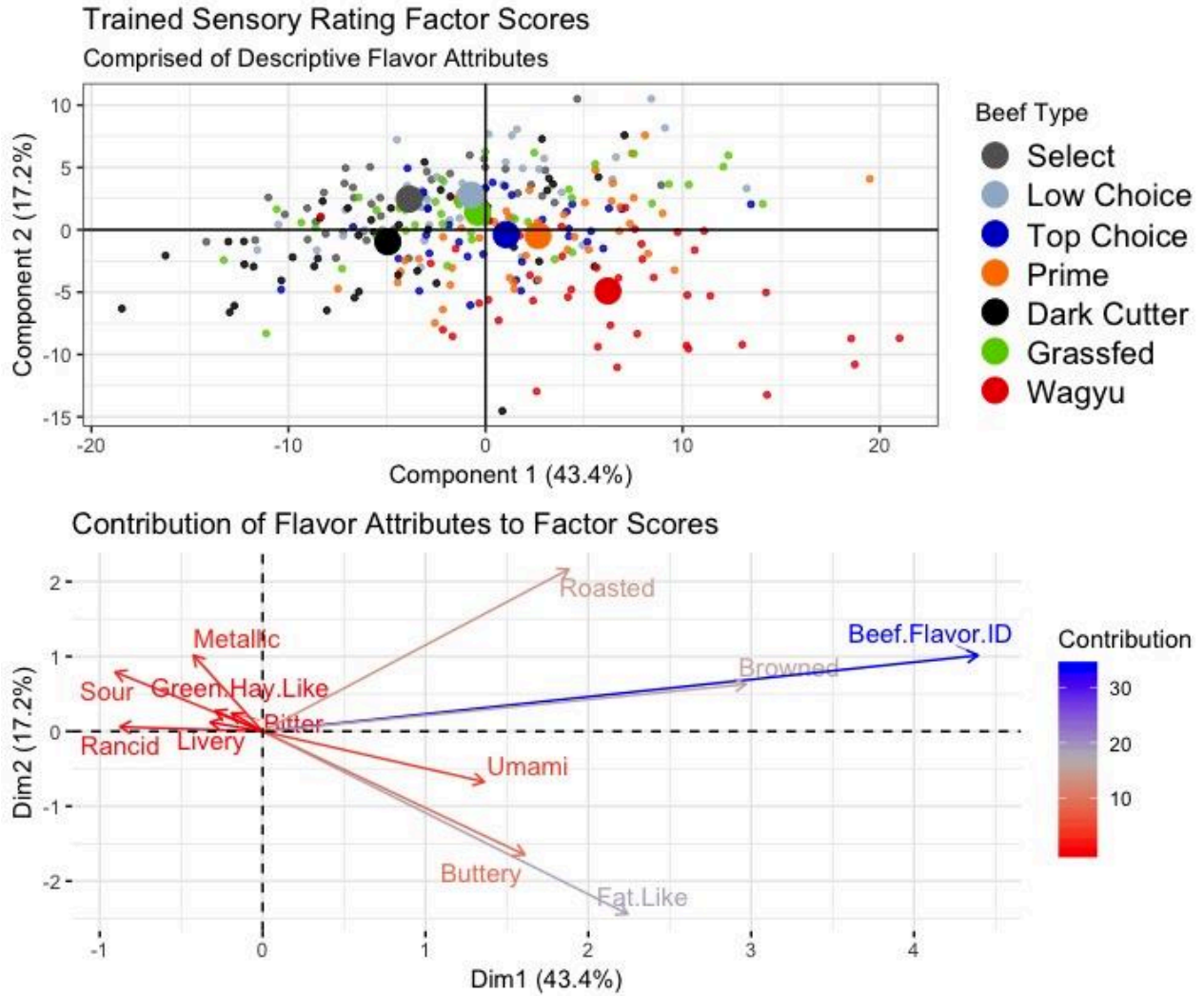
Figure 3. 7 (top) Projection of principal component scores of trained sensory ratings for flavor attributes. Treatment centers are represented by large points. (bottom) Loadings plot showing the contribution of each sensory attribute to factor scores.
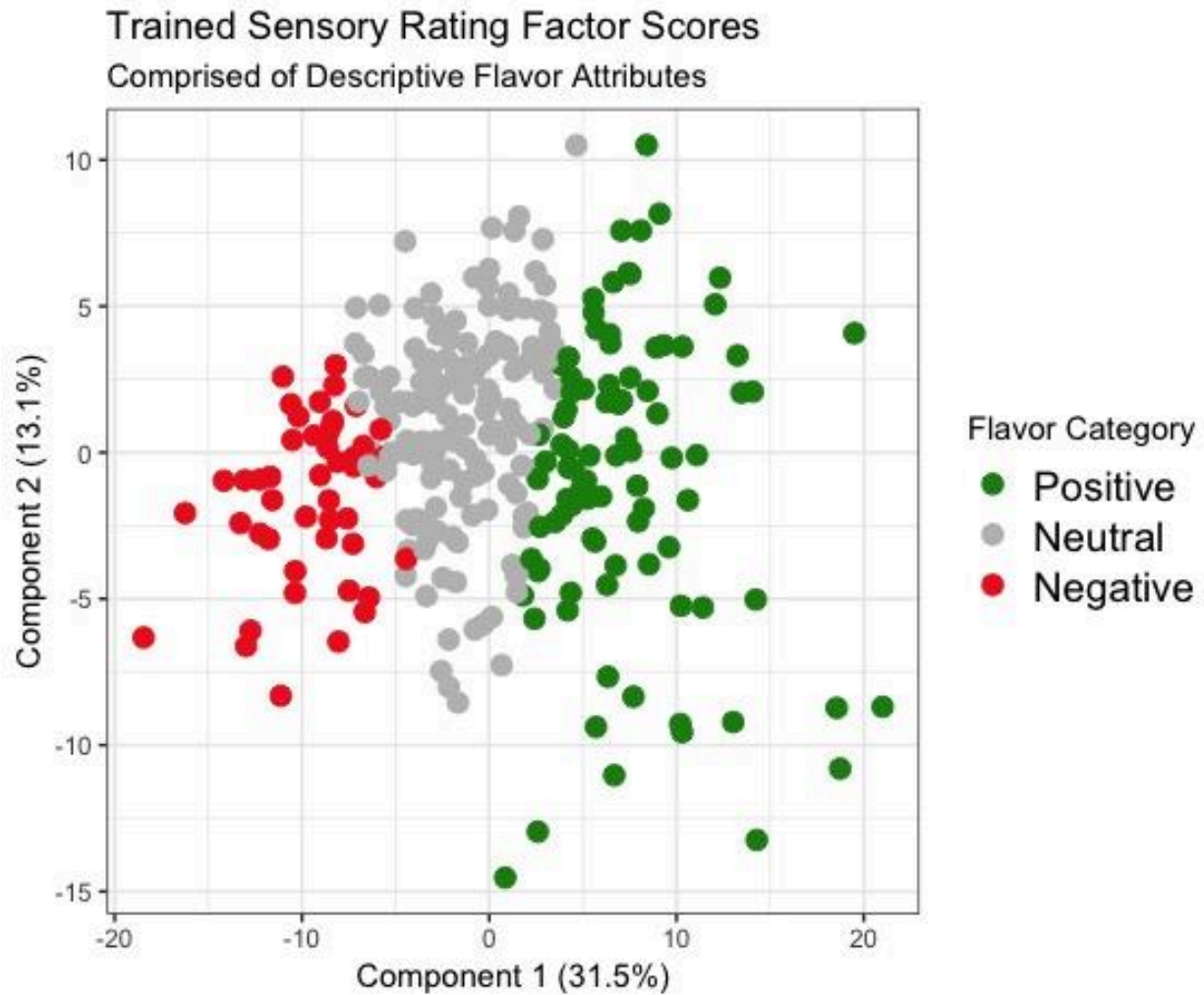
Figure 3. 8 Projection of principal component scores of trained sensory ratings for tenderness, juiciness, and flavor attributes. Scores are colored to represent overall sensory categories as determined by cluster analysis.
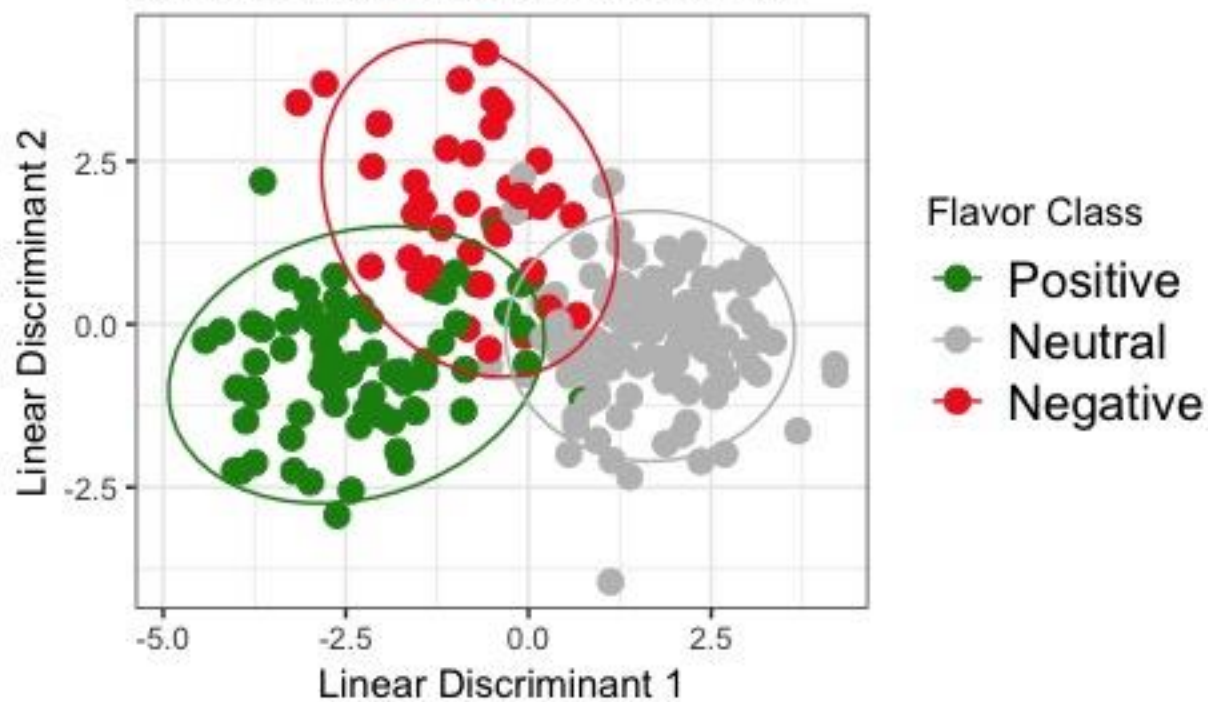
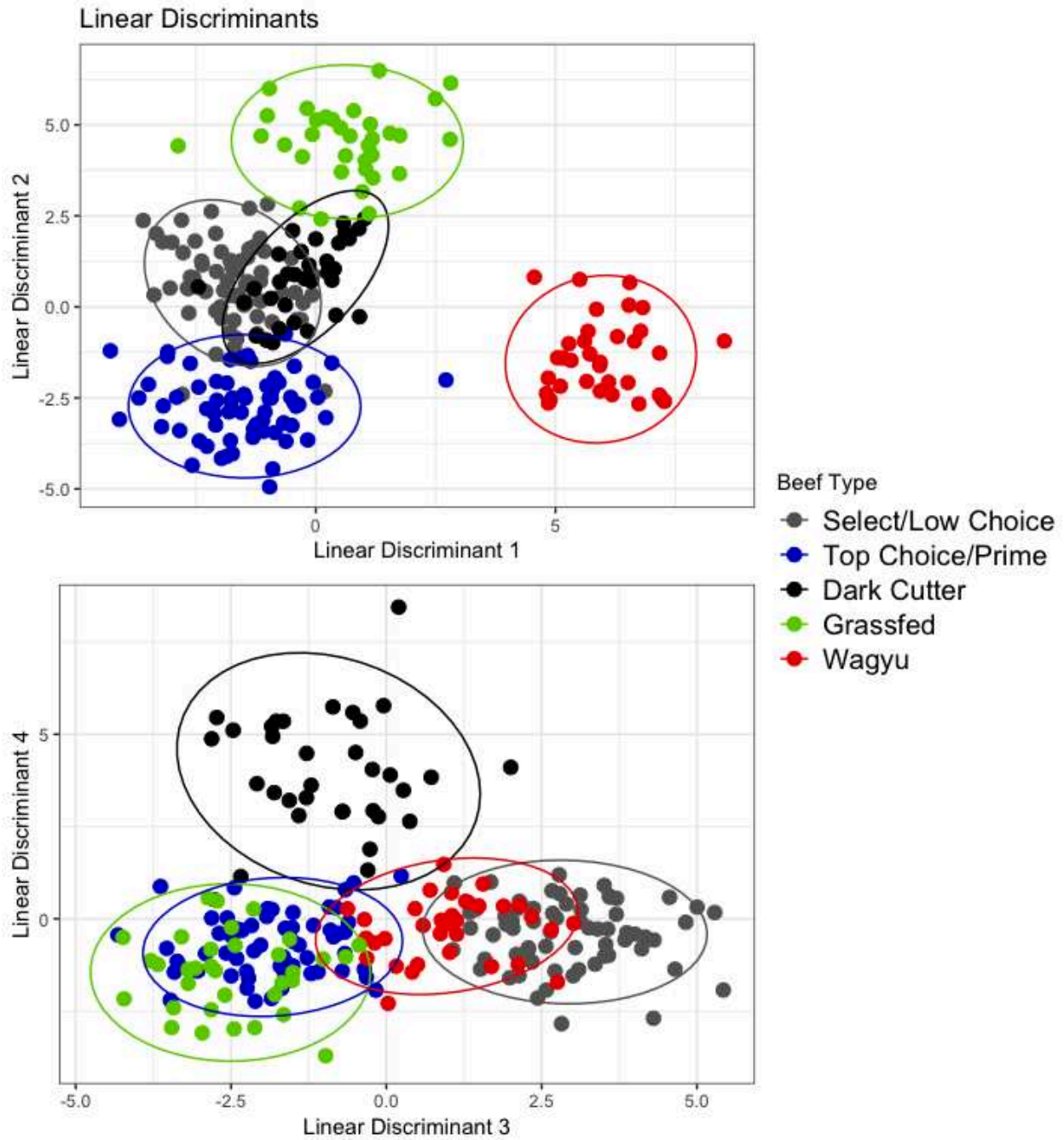**Figure 3. 9** Projection of partial least squares-linear discriminant scores of the training model built from rapid evaporative ionization mass spectrometry (REIMS) mass bins to predict flavor class.
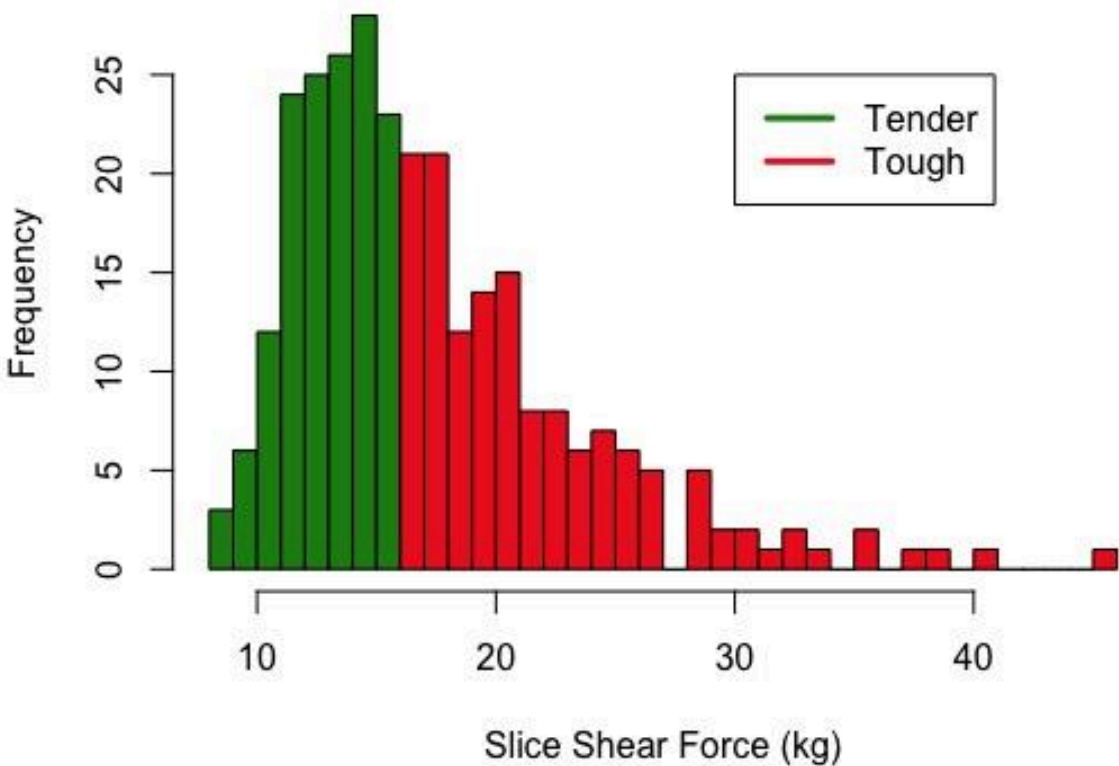
**Figure 3. 10** Distribution of SSF values and assignment of tenderness classifications of beef strip steaks.
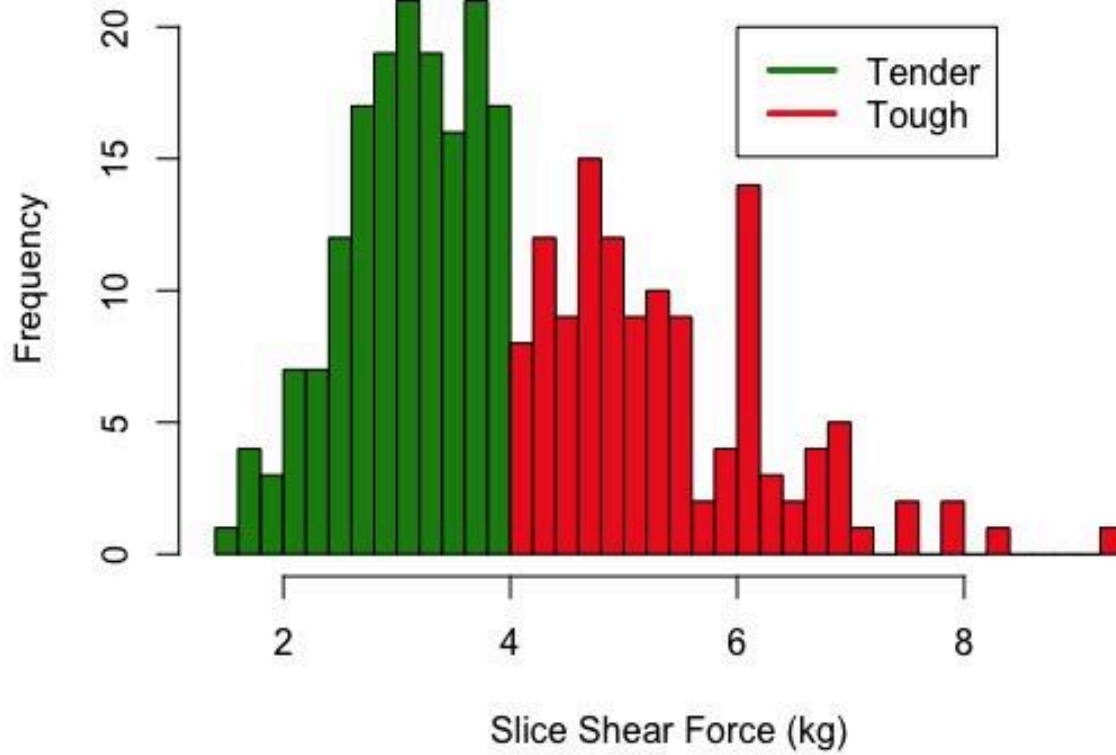
Figure 3. 11 Distribution of WBS values and assignment of tenderness classifications of beef strip steaks.

LITERATURE CITED

Astel, A., S. Tsakovski, P. Barbieri, and V. Simeonov. 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Res. 41:4566–4578.

Balog, J., D. Perenyi, C. Guallar-Hoyas, A. Egri, S. D. Pringle, S. Stead, O. P. Chevallier, C. T. Elliott, and Z. Takats. 2016. Identification of the Species of Origin for Meat Products by Rapid Evaporative Ionization Mass Spectrometry. J. Agric. Food Chem. 64:4793–4800.

Balog, J., T. Szaniszlo, K. Schaefer, J. Denes, A. Lopata, L. Godorhazy, D. Szalay, L. Balogh, L. Sasi-szabo, M. Toth, and Z. Takats. 2010. Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry. Anal. Chem. 82:7343–7350.

Belk, K. E., M. H. George, J. D. Tatum, G. G. Hilton, R. K. Miller, M. Koohmaraie, J. O. Reagan, and G. C. Smith. 2001. Evaluation of the Tendertec beef grading instrument to predict the tenderness of steaks from beef carcasses. J. Anim. Sci. 79:688–697.

Black, C., Olivier, P. Chevallier, S. A. Haughey, J. Balog, S. Stead, S. D. Pringle, M. V Riina, F. Martucci, P. L. Acutis, M. Morris, D. S. Nikolopoulos, Z. Takats, and C. T. Elliott. 2017. A real time metabolomic profiling approach to detecting fish fraud using rapid evaporative ionisation mass spectrometry. Metabolomics 13.

Boleman, S. J., S. L. Boleman, R. K. Miller, J. F. Taylor, H. R. Cross, T. L. Wheeler, M. Koohmaraie, S. D. Shackelford, M. F. Miller, R. L. West, D. D. Johnson, and J. W. Savell. 1997. Consumer Evaluation of Beef of Known Categories of Tenderness. J. Anim. Sci. 75:1521–1524.

Bro, R., and A. K. Smilde. 2014. Principal component analysis. J. Chemometrics. 17:16-33.

Emerson, M. R., D. R. Woerner, K. E. Belk, and J. D. Tatum. 2013. Effectiveness of USDA instrument-based marbling measurements for categorizing beef carcasses according to differences in longissimus muscle sensory attributes. J. Anim. Sci. 91:1024–1034.

Ghatak, A. 2017. Machine Learning with R. Springer, Singapore.

Konda, G., N. Ae, L. M. Grimes, J. Subbiah, C. R. Calkins, A. Samal, and G. E. Meyer. 2008. Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. Sens. Instrum. Food Qual. 2:178-188.

Konda, N. G., K. Cluff, A. Samal, C. R. Calkins, D. D. Jones, G. E. Meyer, and J. Subbiah. 2016. Three dimensional chemometric analyses of hyperspectral images for beef tenderness forecasting. J. Food Eng. 169:309–320.

Kuhn, M., and K. Johnson. 2016. Applied Predictive Modeling. Springer, New York, NY.

Mafi, G., B. Harsh, and J. Scanga. 2014. Review of Instrument Augmented Assessment of USDA Beef Carcass Quality Grades. In: 67th Annual Repriocal Meat Conference.

Mahmood, S., J. A. Basarab, W. T. Dixon, and H. L. Bruce. 2016. Can potential for dark cutting be predicted by phenotype? Relationship between sex, carcass characteristics, and the incidence of dark cutting beef.J. Plaizier, editor. Can. J. Anim. Sci. 96:19–31.

Martinez, H. A., A. N. Arnold, J. C. Brooks, C. C. Carr, K. B. Gehring, D. B. Griffin, D. S. Hale, G. G. Mafi, D. D. Johnson, C. L. Lorenzen, R. J. Maddock, R. K. Miller, D. L. VanOverbeke, B. E. Wasser, and J. W. Savell. 2017. National Beef Tenderness Survey–2015: Palatability and Shear Force Assessments of Retail and Foodservice Beef. Meat Muscle Biol. 1:138–148.

Miller, M. F., M. a Carr, C. B. Ramsey, K. L. Crockett, and L. C. Hoover. 2001. Consumer thresholds for establishing the value of beef tenderness. J. Anim. Sci. 79:3062–3068.

Nubiato, K. E. Z., M. R. Mazon, D. S. Antonelo, C. R. Calkins, G. K. Naganathan, J. Subbiah, and S. da Luz e Silva. 2018. A bench-top hyperspectral imaging system to classify beef from Nellore cattle based on tenderness. Infrared Phys. Technol. 89:247–254.

Park, B., Y. R. Chen, W. R. Hruschka, S. D. Shackelford, and M. Koohmaraie. 1998. Near-infrared reflectance analysis for predicting beef longissimus tenderness. J. Anim. Sci. 76:2115.

Platter, W. J., J. D. Tatum, K. E. Belk, S. R. Koontz, P. L. Chapman, and G. C. Smith. 2013. Effects of marbling and shear force on consumers' willingness to pay for beef strip loin steaks. J. Anim. Sci. 83:890–899.

Price, D. M., G. G. Hilton, D. L. VanOverbeke, and J. B. Morgan. 2008. Using the near-infrared system to sort various beef middle and end muscle cuts into tenderness categories. J. Anim. Sci. 86:413–418.

Shackelford, S. D., T. L. Wheeler, M. Koohmaraie, and R. L. Hruska. 1999. Tenderness Classification of Beef: II. Design and Analysis of a System to Measure Beef Longissimus Shear Force Under Commercial Processing Conditions 1. J. Anim. Sci 77:1474–1481.

Smith, G. C., J. D. Tatum, and K. E. Belk. 2008. International perspective: characterisation of United States Department of Agriculture and Meat Standards Australia systems for assessing beef quality. Aust. J. Exp. Agric. 48:1465.

St John, E. R., J. Balog, J. S. McKenzie, M. Rossi, A. Covington, L. Muirhead, Z. Bodai, F. Rosini, A. V. M. Speller, S. Shousha, R. Ramakrishnan, A. Darzi, Z. Takats, and D. R. Leff. 2017. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. Breast Cancer Res. 19:59.

Verplanken, K., S. Stead, R. Jandova, C. Van Poucke, J. Claereboudt, J. Vanden Bussche, S. De Saeger, Z. Takats, J. Wauters, and L. Vanhaecke. 2017. Rapid evaporative ionization mass

spectrometry for high-throughput screening in food analysis: The case of boar taint. Talanta 169:30–36.

Vote, D. J., K. E. Belk, J. D. Tatum, J. A. Scanga, and G. C. Smith. 2003. Online prediction of beef tenderness using a computer vision system equipped with a BeefCam module. J. Anim. Sci. 81:457–465.

Wheeler, T. L., S. D. Shackelford, and M. Koohmaraie. 2004. The accuracy and repeatability of untrained laboratory consumer panelists in detecting differences in beef longissimus tenderness. J. Anim. Sci. 82:557.

Wheeler, T. L., D. Vote, J. M. Leheska, S. D. Shackelford, and K. E. Belk. 2002. The Efficacy of Three Objective Systems for Identifying Beef Cuts That Can Be Guaranteed Tender.

Woerner, and K. E. Belk. 2008. The History of Instrument Assessment of Beef. Report Submitted to National Cattlemen's Beef Association, Denver, CO.

Xiong, H., J. Wu, and L. Liu. 2010. Classification with Class Overlapping: A Systematic Study. 2010 Int. Conf. E-bus. Intell.:491–497.

REIVIEW OF LITERATURE – PART II

Beef Eating Quality

Although eating quality, in a general sense, is driven by a myriad of factors, beef eating quality most commonly refers to the three attributes of tenderness, flavor, and juiciness. In a survey where consumers were asked to rank attributes that motivated their beef purchasing decisions, tenderness, juiciness, and flavor were ranked the highest among a list of attributes that included other factors such as price, product consistency, and ease of preparation (Reicks et al., 2011). With consumer satisfaction being closely related to beef eating quality, significant research emphasis has been placed on understanding the intrinsic and extrinsic factors that influence the development of tenderness, juiciness, and flavor. The importance of eating quality and consumer satisfaction is clearly evident in the way that beef is merchandised in the United States. Most beef carcasses are merchandised using a grid-based system, taking USDA yield and quality grades into consideration when determining overall carcass value. The USDA beef grading standards define quality grade as "the characteristics of the meat which predict the eating quality of the lean" (USDA, 2017). Higher quality grades receiving premiums, as they are associated with steaks and roasts that are more tender, more flavorful, and juicier, emphasizing the importance placed on eating quality for the merchandising of beef products.

In addition to understanding these attributes individually, it is important to understand how these attributes interact to influence consumer overall acceptability of beef products. Historically, tenderness has been considered the most important eating quality trait when determining a consumer's overall acceptability of beef (G. Smith et al., 1987; Miller et al.,

2001). However; if tenderness is acceptable, then flavor has the greatest relationship to overall acceptability (Corbin et al., 2014). Beef tenderness has greatly improved throughout the years and there are several muscles that rarely fall below the tenderness threshold when production parameters are correctly managed (Martinez et al., 2017). Yet, inconsistencies remain, and several muscles suffer from tenderness issues. Therefore, although flavor has become more influential in its contribution to consumer acceptability, tenderness continues to be of great interest and continues to be heavily researched. In a recent meta-analysis of factors influencing the overall consumer beef eating experience, it was concluded that the lack of acceptability of even a single eating quality trait significantly decreases the probability of a positive overall experience (O'Quinn et al., 2018). Thus, neither tenderness, flavor, nor juiciness can be disregarded when evaluating factors affecting eating quality development.

Beef Tenderness

Of all eating quality traits, tenderness has received the greatest research emphasis throughout the years (Aberle et al., 2001). Several authors have concluded that consumer ratings for tenderness have the strongest relationships to overall eating quality (Shackelford et al., 2001). Along with this relationship between tenderness and overall acceptability, consumers have indicated an increased willingness to pay for tender product (Boleman et al., 1997; Miller et al., 2001; Shackelford et al., 2001). Platter et al. (2013) evaluated consumer willingness to pay for beef with varying marbling scores and tenderness levels in a sealed-bid auction scenario. Consumers were recruited to rate sensory attributes of beef strip steaks and were given $40 compensation for their time. After each sensory session, each panelist was given the opportunity to participate in an auction to bid on steaks identical to those evaluated in the panel using their compensation money. Consumers were given a base price of $14.32/kg as an average retail price

as a starting point for their bids, with the ability to bid higher or lower. Consumers were able to detect differences in tenderness based on WBS values and, on average, decreased their bids by $1.02/kg for each 1 kg increase in WBS values (Platter et al., 2013).

The development and perception of beef tenderness is influenced by numerous factors, both intrinsically and extrinsically. From a simplistic viewpoint, beef tenderness is affected by three muscle characteristics: collagen content, sarcomere length, and structure of myofibrillar proteins (Koohmaraie et al., 2002). However, when one begins to consider the factors influencing each of these three components, the true complexity of tenderness comes to light. Collagen content can be influenced by factors such as breed type, animal maturity, and muscle, among others (Purslow, 2005). Sarcomere length is also largely influenced by muscle, but is heavily reliant on postmortem metabolic processes and the temperature at which these pathways proceed (King et al., 2003; Tschirhart-Hoelscher et al., 2006).

In addition to several of the factors previously mentioned, the structure of myofibrillar proteins and its relationship to beef tenderness is significantly related to the extent of postmortem aging (Koohmaraie and Geesink, 2006). The aforementioned characteristics relating to beef tenderness are dependent on the physiological make-up of the live animal and the metabolic processes occurring during the harvest process and the postmortem period but is in no way an extensive list of characteristics influencing beef tenderness. Even after the conversion of muscle to meat, tenderness can still be manipulated through several extrinsic preparation methods. Furthermore, regardless of the muscle components of beef, tenderness is significantly manipulated during the cooking process as a result of cooking method, cooking temperature, and final degree of doneness (Parrish et al., 1973; Belk et al., 1993; Yancey et al., 2011). Although

this is not an exhaustive list of beef tenderness determinants, it begins to show the complexity and multi-faceted nature of beef tenderness.

Collagen

Collagen is the major component of intramuscular connective tissue (IMCT; Lepetit, 2008) and contributes to what is known as "background toughness" of beef. This term was originally coined because it is a source of toughness that occurs at the time of slaughter but does not change much during the slaughter process or the storage period. Collagen content will vary from muscle-to-muscle and among animals of different maturity levels (Hill, 1966; Jeremiah et al., 2003); however, if factors such as muscle and animal age are held constant, collagen content does little to explain animal-to-animal tenderness variation. Nevertheless, increased toughness associated with increased collagen content is reflected in price differences among various beef muscles seen at retail (Purslow, 2005). Collagen is found in three layers within muscle: the epimysium, perimysium, and endomysium (McCormick, 1994). The epimysium is the outer most connective tissue layer surrounding the entire muscle. Particularly in single muscle beef cuts, the epimysium has a lesser contribution to beef tenderness relative to the other connective tissue layers because it is many times removed before cooking and consumption (Purslow, 2014).

Perimysium is the intermediate layer of connective tissue surrounding individual muscle bundles and is responsible for visual differences in muscle texture, accounting for as much as 90% of total connective tissue in muscle (McCormick, 1999). Of all connective tissue layers, it is the most variable from muscle-to-muscle and is believed to contribute the greatest to tenderness differences (Purslow, 2005). When several beef muscles were aged for 14 d, Brooks and Savell (2004) found perimysium thickness alone to explain 20% of variation in WBS values. Finally,

the endomysium is the inner most layer of connective tissue surrounding individual muscle fibers.

The relationship between collagen content and beef tenderness appears to be quite variable and dependent on the combination of numerous factors. Previously published literature has been inconsistent in establishing a clear relationship between collagen content and tenderness, with some reporting strong relationships (Torrescano et al., 2003; Riley et al., 2005), while others report collagen to explain little to no variation in beef tenderness (Cross et al., 1973; Seideman, 1986; Campo et al., 2000; Serra et al., 2008). In the cases of the latter studies, relationships were only evaluated within a single muscle. Seideman et al. (1987) evaluated various factors influencing tenderness from beef striploins and concluded the majority of tenderness variation could be explained by the myofibrillar component. These results further supported the idea that connective tissue contributed to background toughness when parameters are held relatively constant, such that, connective tissue was more than likely contributing to the overall toughness of the samples but did not vary enough from animal to animal to explain tenderness variation. However, Torrescano et al. (2003) evaluated the relationship among 14 different beef muscles and found large collagen content differences among muscles, as well as, strong relationships to WBS values, supporting the notion that collagen content better explains tenderness variation between different muscles.

There is a common belief that as the total amount of collagen increases, beef becomes tougher. Although total collagen content only slightly varies within the same muscle of different animals, total collagen content does greatly vary from muscle-to-muscle. As mentioned above, other muscle components largely influence beef tenderness; thus, a high total collagen content does not necessarily equate to an excessively tough muscle. Because this is generally the case, it

has led many to develop an overly simplistic idea of the relationship between total collagen content and beef tenderness. For instance, Torrescano et al. (2003) found beef *infraspinatus* and *pectoralis profundus* muscles to have similarly high amounts of total collagen content. Throughout the entirety of the study, total collagen content was found to be highly correlated (r = 0.723) to Warner-Bratlzer shear force (WBS) values. However, *infraspinatus* produced an average WBS value that was nearly 2.5 kg lower than *pectoralis profundus*. This observation clearly indicates that other components counterbalanced the high collagen content of *infraspinatus* to result in a moderately tender muscle. There is a complicated relationship among factors influencing tenderness development, and whether a muscle is perceived as tough or tender depends on a specific combination of multiple characteristics.

Not only does the amount of collagen influence tenderness, but the extent of cross-linking also has a substantial impact on beef tenderness. Without cross-links, collagen would have very minimal structural qualities (Bailey and Light, 1989). When a collagen fiber is thermally denatured, it shrinks, pulling attached muscles inward, creating tension (McCormick, 1999). The degree to which collagen fibers shrinks is a reflection of the thermal stability of those crosslinks (McCormick, 1999). In terms of thermal stability, collagen exists in heat soluble and heat insoluble forms, with the majority of total collagen being heat insoluble, or heat stable (Gredell et al., 2018). When denaturation temperatures are reached, heat soluble collagen gelatinizes, as opposed to shrinking, and is removed with the exudate. Because it is removed from the system, heat soluble collagen is believed to have a much lesser influence on tenderness development. Thus, as the percent of heat stable collagen increases, an increase in toughness is generally observed (Hill, 1966). Not only does the thermal stability of collagen increase as an animal

matures (Gredell et al., 2018), but the ratio of heat stable to heat soluble collagen crosslinks is also increased in tougher muscles (Light et al., 1985).

Thermal Properties of Collagen

The composition of collagen and its thermal properties play a substantial role in determining cooked meat texture and perceived tenderness. Thus, an ample amount of research emphasis has been placed on understanding the complexities of collagen thermodynamics. Textural characteristics of collagen are clearly altered during the cooking process. Relationships between collagen characteristics and raw meat texture are generally very strong; however, when cooked, that relationship can be completely lost (M. Christensen et al., 2011). When exposed to high enough temperatures, collagen will begin to denature. Collagen is unique in that, as it is thermally denatured, it shrinks and produces tension (Du and McCormick, 2009). Denaturation temperature is highly variable from species to species, and is largely dependent on the extent of post-transitional hydroxylation of the amino acid proline to hydroxyproline (Bailey and Light, 1989). The addition of a hydroxyl group to proline facilitates the formation of hydrogen bonds between adjacent molecules, adding to the structural properties of collagen (Alexander Rich and Crick, 1955). Therefore, as hydroxyproline content increases, so does the thermal denaturation temperature of collagen (Bailey and Light, 1989).

In meat tissues, collagen generally shrinks around 65°C (Lepetit, 2008) to one-quarter of its length if left unrestrained (Tornberg, 2005). As the amount of heat stable crosslinks increases; however, so does the thermal shrinkage temperature (Smith and Judge, 1991). Due to this shrinkage and tension development, collagen is mainly responsible for the decrease in tenderness observed when meat is cooked between 65 and 75°C (Davey and Gilbert, 1974). But, if the crosslinks are heat soluble aldimine crosslinks, collagen will dissolve and gelatinize, resulting in

an observed decrease in tenderness. Even when exposed to severely high temperatures, heat stable crosslinks will resist gelatinization, remain intact, and produce a rubber-like texture (Bailey and Light, 1989). For this reason, it can be argued that the total amount of heat stable collagen, or even the ratio of heat stable to heat insoluble collage, may be a better indication of collagen's contribution to cooked meat texture. McCormick (1999) suggested that the amount of heat stable collagen and the total amount of collagen have an additive effect on meat toughness. As mentioned previously, though, there are numerous muscle components influencing tenderness development and the strength of the relationship between collagen and toughness appears to be relative to other endogenous and exogenous factors. Correlations between collagen content and tenderness are more prevalent when comparing various muscles (Dransfield, 1977; Ngapo et al., 2002), but are much lower when muscle source is held constant (Cross et al., 1973; Miller et al., 1983; Serra et al., 2008).

Cooking Rate

During the cooking process, toughening occurs in a bi-phasic manor, the first between 40 and 50°C and the second between 65 and 75°C (Davey and Niederer, 1977). The first toughening phase is due to the denaturation and coagulation of myofibrillar proteins, whereas, the second toughening phase is due to the denaturation and shrinkage of collagen (Bailey and Light, 1989). A third increase in toughness has even been reported to occur upwards of 90°C, associated with the denaturation and shrinkage of actin (Bailey and Light, 1989; Palka and Daun, 1999). At 45-60°C muscle fiber shrinkage stems from a decrease in diameter, whereas, at 60-90°C, shrinkage primarily occurs longitudinally (Palka and Daun, 1999). When more rapid cooking methods are used, an increase in degree of doneness results in a tougher product. Although collagen gelatinizes at temperatures associated with higher degrees of doneness, it is likely that the rapid

cooking process does not hold collagen to temperatures above those needed for denaturation to allow for gelatinization. Additionally, moisture lost due to the shrinkage of collagen and compression of muscle fibers also adds to the toughness of meat cooked to higher degrees of doneness. Martens et al. (1982) reported an increase in sensory firmness with the denaturation of myosin and actin at 40-60°C and 66-72°C, respectively, and a decrease in fiber cohesivity at 62°C. Panelists then determined optimal texture preference to be in the range of 60-67°C, which would indicate that myosin and collagen would be denatured, but myosin would remain intact.

Vasanthi et al. (2007) found shear force and sensory tenderness to improve as both water bath temperature and cooking time increased from 80-100°C and 30-60 min, respectively. The authors also measured an increase in collagen solubility, as well as, a decrease in total collagen. It was suggested that increase cooking time and temperature resulted in further gelatinization of collagen, which was released with the cooking juice (Vasanthi et al., 2007). Christensen et al. (2013) also determined an increase in cooking temperature and time to increase beef *semitendinosus* from young bulls and cows. Additionally, beef from cows required an extended cooking time to reach the same tenderness level as beef rom young bulls. Yet, even after cooking for 19.5 h, young bulls still had greater collagen solubility. Interestingly, heat soluble collagen from young bulls was affected by cooking time and not cooking temperature, whereas, heat soluble collagen from cows was affected by cooking temperature and not cooking time (Christensen et al., 2013). This likely demonstrated increased collagen thermal stability of mature beef due to an increase in collagen crosslinking. Additionally, activity of cathepsins B and L were monitored throughout the cooking process. Activity decreased with cooking temperature and time; however, cathepsin activity remained in beef cooked at 53°C even after 7.5 h of cooking.

In addition to collagen solubilization, cooking meat at low temperatures for extended periods of time provides tenderness improvement through continued proteolytic enzyme activity (Tornberg, 2005). Cathepsins are a group of enzymes capable of degrading several myofibrillar proteins and collagen (Beltrán et al., 1992; Nowak, 2011). Their role in postmortem proteolysis has been extensively researched, but the general consensus is that they do not contribute to the aging response because they are stored in the lysosomes and do not get released postmortem (Du and McCormick, 2009). However, cathepsins have been collected from cook loss and have shown to retain activity during cooking up to 63°C (L. Christensen et al., 2011). Since cathepsins are clearly released during cooking, it has been suggested that the heating process disrupts the lysosome wall, allowing for the release of cathepsins and enzymatic action during the cooking process. While important to tenderness development during postmortem aging, calpain does not contribute to tenderness improvements seen with slow cooking rates, as it quickly loses functionality at 55°C (Ertbjerg et al., 2012). Furthermore, L. Christensen et al. (2011) showed a negative relationship ($r = -0.50$) between cathepsin activity and shear force values. Penfield and Meyer (1975) did not target specific enzymes but found enzymes to be active during the heating of meat. Their results showed enzymatic activity was greatest between 50-60°C, which was also where the greatest decrease in shear force values occurred (Penfield and Meyer, 1975). Knowing cathepsins have the ability to degrade myofibrillar proteins and collagen, are released from the lysosomes during cooking, and retain proteolytic activity up to 63°C suggests a strong influence on tenderness improvement during low temperature, long time cooking treatments.

Humidity in the Cooking Environment

Recently, popularity of combination steam ovens have increased with both foodservice and in-home applications. Combination steam ovens are able to provide users the ability to cook

with pure steam, a combination of hot air and steam, or as a traditional convection oven. Manufacturers promote the use of combination ovens through increased cooking efficiencies, increased product yields, precise control over the cooking environment, and increased health claims, among others. According to recent market research, demand for combination ovens is expected to increase by 25% globally over the next five years (ARC, 2018). With the rise in popularity of meal-delivery kits, combination steam ovens are even finding their way into meal-delivery kits by utilizing countertop combination ovens with custom cooking cycles to fit the needs of each individual meal. If these trends continue to rise, then understanding the influence of added humidity during the cooking process on meat quality and eating quality would allow for optimum utilization of this technology.

Moist-heat cookery methods are recommended for the cooking of inherently tougher muscles. Early studies have concluded that moist-heat cookery methods aide in the degradation of collagen during cooking, resulting in a more tender product (Cover and Smith, 1956). Although little to no work has been conducted to specifically evaluate the influence of humidity on the thermal characteristics of collagen in meat, previous studies have shown that lower water content increases denaturation temperature and gelatinization of collagen in leather (Badea et al., 2012). Furthermore, recent correlations have been made between an increase in collagen denaturation temperature and a decrease in water content in an effort to manipulate the stability of collagen in synthetic non-meat applications (Fratzl, 2008). Without an excess of moisture in the collagen fiber, it is believed the cross-links are able to form a tighter bond. The addition of moisture to the cooking environment has been attributed to an increase in tenderness; however has yet to be related to collagen solubility (Kolle et al., 2004). In contrast, other studies have shown the opposite; that cooking with humidity decreases tenderness (Berry et al., 1977;

94

Chiavaro et al., 2009; Isleroglu et al., 2015). Due to water's greater capacity for heat transfer, adding humidity to the cooking environment increases cooking rate, as well as, cooking yields (Isleroglu et al., 2014). Increasing cooking rate of dry-heat methods increases beef toughness (King et al., 2003); therefore, an increase cooking rate more than likely explains toughening of meat cooked in the presence of added humidity. Particularly in inherently tougher muscles; however, utilizing steam in the cooking environment could be beneficial if oven temperature is lowered to adjust for the increased cooking rate from humidity so that advantages in tenderness improvement from slower cooking rates can be achieved, while controlling for excess moisture loss.

LITERATURE CITED

Aberle, E. D., J. C. Forrest, D. E. Gerrard, and E. W. Mills. 2001. Principles of Meat Science. 4th ed. Kendall/Hunt Publishing Company.

Alexander Rich, D., and D. F. H. C. Crick. 1955. The Structure of Collagen. Nature 175:915-916.

Badea, E., G. Della Gatta, and T. Usacheva. 2012. Effects of temperature and relative humidity on fibrillar collagen in parchment: A micro differential scanning calorimetry (micro DSC) study. Polym. Degrad. Stab. 97:346–353.

Bailey, A. J., and N. D. Light. 1989. Connective Tissue In Meat and Meat Products. Elsevier Applied Science.

Belk, K. E., G. L. Luchak, and R. K. Miller. 1993. Physical Characterisitics of Beef Roasts Prepared with Different Foodservice Cooking Methods. J. Muscle Foods 4:119–139.

Beltrán, J., M. Bonnet, and A. Ouali. 1992. Comparative action of cathepsins B and L on intramuscular collagen as assessed by differential scanning calorimetry. Meat Sci. 32:299–306.

Berry, B. W., M. R. Wheeling, and J. A. Carpenter. 1977. Effects of muscle and cookery method on palatability of beef from several breeds and breed crosses. J. Food Sci. 42:1322–1324.

Boleman, S. J., S. L. Boleman, R. K. Miller, J. F. Taylor, H. R. Cross, T. L. Wheeler, M. Koohmaraie, S. D. Shackelford, M. F. Miller, R. L. West, D. D. Johnson, and J. W. Savell. 1997. Consumer Evaluation of Beef of Known Categories of Tenderness. J. Anim. Sci. 75:1521–1524.

Campo, M. M., P. Santolaria, C. Sañudo, J. Lepetit, J. L. Olleta, B. Panea, and P. Albertí. 2000. Assessment of breed type and ageing time effects on beef meat quality using two different texture devices. Meat Sci. 55:371–8.

Chiavaro, E., M. Rinaldi, E. Vittadini, and D. Barbanti. 2009. Cooking of pork Longissimus dorsi at different temperature and relative humidity values: Effects on selected physico-chemical properties. J. Food Eng. 93:158–165.

Christensen, L., P. Ertbjerg, M. D. Aaslyng, and M. Christensen. 2011. Effect of prolonged heat treatment from 48 °C to 63 °C on toughness, cooking loss and color of pork. Meat Sci. 88:280–5.

Christensen, L., P. Ertbjerg, H. Løje, J. Risbo, F. W. J. van den Berg, and M. Christensen. 2013. Relationship between meat toughness and properties of connective tissue from cows and young bulls heat treated at low temperatures for prolonged times. Meat Sci. 93:787–95.

Christensen, M., P. Ertbjerg, S. Failla, C. Sañudo, R. I. Richardson, G. R. Nute, J. L. Olleta, B. Panea, P. Albertí, M. Juárez, J.-F. Hocquette, and J. L. Williams. 2011. Relationship between collagen characteristics, lipid content and raw and cooked texture of meat from young bulls of fifteen European breeds. Meat Sci. 87:61–65.

Corbin, C. H., T. G. O'Quinn, A. J. Garmyn, J. F. Legako, M. R. Hunt, T. T. N. Dinh, R. J. Rathmann, J. C. Brooks, and M. F. Miller. 2014. Sensory evaluation of tender beef strip loin steaks of varying marbling levels and quality treatments. Meat Sci. 100:24–31.

Cover, S., and W. H. Smith. 1956. The effect of two methods of cooking on palatability scores, shear force values, and collagen content of two cuts of beef. J. Food Sci. 21:312–321.

Cross, H. R., Z. L. Carpenter, and G. C. Smith. 1973. Effect of Intramuscular Collagen and Elastin on Bovine Muscle Tenderness. J. Food Sci. 38:998–1003.

Davey, C. L., and K. V. Gilbert. 1974. Temperature-dependent cooking toughness in beef. J. Sci. Food Agric. 25:931–938.

Davey, C. L., and A. F. Niederer. 1977. Cooking tenderizing in beef. Meat Sci. 1:271–6.

Dransfield, E. 1977. Intramuscular composition and texture of beef muscles. J. Sci. Food Agric. 28:833–842.

Du, M., and R. J. McCormick. 2009. Applied Muscle Biology and Meat Science. First. CRC Press, Boca Raton, FL.

Ertbjerg, P., L. S. Christiansen, A. B. Pedersen, and L. Kristensen. 2012. The effect of temperature and time on activity of calpains and lysosomal enzymes and degradation of desmin in porcine longissimus muscle. In: 58th International Congress of Meat Science and Technology. p. 12–17.

Fratzl, P. 2008. Collagen Structure and Mechanics. In: Springer, New York, NY.

Gredell, D. A., T. G. O'Quinn, J. F. Legako, J. C. Brooks, and M. F. Miller. 2018. Palatability and Biochemical Factors of Beef from Mature Cattle Finished on a Concentrate Diet Prior to Harvest. Meat Muscle Biol. 2:111.

Hill, F. 1966. The solubility of intramuscular collagen in meat animals of various ages. J. Food Sci. 31:161–166.

Isleroglu, H., T. Kemerli, and F. Kaymak-Ertekin. 2015. Effect of steam-assisted hybrid cooking on textural quality characteristics, cooking loss, and free moisture content of beef. Int. J. Food Prop. 18:403–414.

Isleroglu, H., T. Kemerli, Ö. Özdestan, A. Uren, and F. Kaymak-Ertekin. 2014. Effect of oven cooking method on formation of heterocyclic amines and quality characteristics of chicken patties: steam-assisted hybrid oven versus convection ovens. Poult. Sci. 93:2296–303.

Jeremiah, L. E., M. E. R. Dugan, J. L. Aalhus, and L. L. Gibson. 2003. Assessment of the chemical and cooking properties of the major beef muscles and muscle groups. Meat Sci. 65:985–992.

King, D. A., M. E. Dikeman, T. L. Wheeler, C. L. Kastner, M. Koohmaraie, D. a. King, M. E. Dikeman, T. L. Wheeler, C. L. Kastner, and M. Koohmaraie. 2003. Chilling and cooking rate effects on some myofibrillar determinants of tenderness of beef. J. Anim. Sci. 81:1473–1481.

Kolle, B. K., D. R. McKenna, and J. W. Savell. 2004. Methods to increase tenderness of individual muscles from beef rounds when cooked with dry or moist heat. Meat Sci. 68:145–54.

Koohmaraie, M., and G. H. Geesink. 2006. Contribution of postmortem muscle biochemistry to the delivery of consistent meat quality with particular focus on the calpain system. Meat Sci. 74:34–43.

Koohmaraie, M., M. P. Kent, S. D. Shackelford, E. Veiseth, and T. L. Wheeler. 2002. Meat tenderness and muscle growth: is there any relationship? Meat Sci. 62:345–352.

Lepetit, J. 2008. Collagen contribution to meat toughness: Theoretical aspects. Meat Sci. 80:960–7.

Light, N., A. E. Champion, C. Voyle, and A. J. Bailey. 1985. The role of epimysial, perimysial and endomysial collagen in determining texture in six bovine muscles. Meat Sci. 13:137–149.

Martinez, H. A., A. N. Arnold, J. C. Brooks, C. C. Carr, K. B. Gehring, D. B. Griffin, D. S. Hale, G. G. Mafi, D. D. Johnson, C. L. Lorenzen, R. J. Maddock, R. K. Miller, D. L. VanOverbeke, B. E. Wasser, and J. W. Savell. 2017. National Beef Tenderness Survey–2015: Palatability and Shear Force Assessments of Retail and Foodservice Beef. Meat Muscle Biol. 1:138–148.

McCormick, R. J. 1994. The flexibility of the collagen compartment of muscle. Meat Sci. 36:79–91.

McCormick, R. J. 1999. Extracellular modifications to muscle collagen: implications for meat quality. Poult. Sci. 78:785–791.

Miller, M. F., M. A. Carr, C. B. Ramsey, K. L. Crockett, and L. C. Hoover. 2001. Consumer thresholds for establishing the value of beef tenderness. J. Anim. Sci. 79:3062–3068.

Miller, R. K., J. D. Tatum, H. R. Cross, R. a Bowling, and R. P. Clayton. 1983. Effects of Carcass Maturity on Collagen Solubility and Palatability of Beef from Grain-Finished Steers. J. Food Sci. 48:484–486.

Ngapo, T.., P. Berge, J. Culioli, E. Dransfield, S. De Smet, and E. Claeys. 2002. Perimysial collagen crosslinking and meat tenderness in Belgian Blue double-muscled cattle. Meat Sci. 61:91–102.

Nowak, D. 2011. Enzymes in tenderization of meat - the system of calpains and other systems - a review. Polish J. Food Nutr. Sci.

O'Quinn, T. G., J. F. Legako, J. C. Brooks, and M. F. Miller. 2018. Evaluation of the contribution of tenderness, juiciness, and flavor to the overall consumer beef eating experience. Transl. Anim. Sci.

Palka, K., and H. Daun. 1999. Changes in texture, cooking losses, and myofibrillar structure of bovine M. semitendinosus during heating. Meat Sci. 51:237–243.

Parrish, F. C., D. G. Olson, B. E. Miner, and R. E. Rust. 1973. Effect of Degree of Marbling and Internal Temperature of Doneness on Beef Rib Steaks. J. Anim. Sci. 37:430.

Penfield, M. P., and B. H. Meyer. 1975. Changes in Tenderness and Collagen of Beef Semitendinosus Muscle Heated at Two Rates. J. Food Sci. 40:150–154.

Platter, W. J., J. D. Tatum, K. E. Belk, S. R. Koontz, P. L. Chapman, and G. C. Smith. 2013. Effects of marbling and shear force on consumers' willingness to pay for beef strip loin steaks. J. Anim. Sci. 83:890–899.

Purslow, P. P. 2005. Intramuscular connective tissue and its role in meat quality. Meat Sci 70:435–447.

Purslow, P. P. 2014. New Developments on the Role of Intramuscular Connective Tissue in Meat Toughness. Annu. Rev. Food Sci. Technol. 5:133–153.

Reicks, A. L., J. C. Brooks, A. J. Garmyn, L. D. Thompson, C. L. Lyford, and M. F. Miller. 2011. Demographics and beef preferences affect consumer motivation for purchasing fresh beef steaks and roasts. Meat Sci. 87:403–11.

Riley, D. G., D. D. Johnson, C. C. Chase, R. L. West, S. W. Coleman, T. A. Olson, and A. C. Hammond. 2005. Factors influencing tenderness in steaks from Brahman cattle. Meat Sci. 70:347–56.

Seideman, S. C. C. 1986. Methods of Expressing Collagen Characteristics and Their Relationship to Meat Tenderness and Muscle Fiber Types. J. Food Sci. 51:273–276.

Serra, X., L. Guerrero, M. D. Guàrdia, M. Gil, C. Sañudo, B. Panea, M. M. Campo, J. L. Olleta, M. D. García-Cachán, J. Piedrafita, and M. A. Oliver. 2008. Eating quality of young bulls from three Spanish beef breed-production systems and its relationships with chemical and instrumental meat quality. Meat Sci. 79:98–104.

Shackelford, S. D., T. L. Wheeler, M. K. Meade, J. O. Reagan, B. L. Byrnes, and M. Koohmaraie. 2001. Consumer impressions of Tender Select beef. J. Anim. Sci. 79:2605.

Smith, G., J. Savell, H. Cross, Z. Carpenter, C. Murphey, G. Davis, H. Abraham, F. Parrish Jr, and B. Berry. 1987. Relationship of USDA Quality Grades To Palatability of Cooked Beef. J. Food Qual. 10:269–286.

Smith, S. H., and M. D. Judge. 1991. Relationship between pyridinoline concentration and thermal stability of bovine intramuscular collagen. J. Anim. Sci. 69:1989–1993.

Tornberg, E. 2005. Effects of heat on meat proteins - Implications on structure and quality of meat products. Meat Sci. 70:493–508.

Torrescano, G., A. Sánchez-Escalante, B. Giménez, P. Roncalés, and J. A. Beltrán. 2003. Shear values of raw samples of 14 bovine muscles and their relation to muscle collagen characteristics. Meat Sci. 64:85–91.

Tschirhart-Hoelscher, T. E., B. E. Baird, D. A. King, D. R. McKenna, and J. W. Savell. 2006. Physical, chemical, and histological characteristics of 18 lamb muscles. Meat Sci. 73:48–54.

Vasanthi, C., V. Venkataramanujam, and K. Dushyanthan. 2007. Effect of cooking temperature and time on the physico-chemical, histological and sensory properties of female carabeef (buffalo) meat. Meat Sci. 76:274–80.

Yancey, J. W. S., M. D. Wharton, and J. K. Apple. 2011. Cookery method and end-point temperature can affect the Warner-Bratzler shear force, cooking loss, and internal cooked color of beef longissimus steaks. Meat Sci. 88:1–7.

CHAPTER V

UNDERSTANDING THE IMPACT OF OVEN TEMPERATURE AND RELATIVE

HUMIDITY ON THE BEEF COOKING PROCESS

Introduction

Tenderness is one of the most important attributes when determining consumer

acceptability of beef (O'Quinn et al., 2012), which was shown to be influenced by cooking

method (Yancey et al., 2011). Therefore, it is critical to establish cooking parameters that

maximize eating quality, without sacrificing efficiency and practicality of the cooking process. In

previous tenderness studies, researchers have accredited the addition of humidity to the cooking

environment as a way to improve the process of tenderization (Kolle et al., 2004; Bowers et al.,

2012). Moisture has shown to be useful in the breakdown of protein and specifically the

solubilization of collagen, which is especially beneficial in the cooking of tougher muscles

(Cover and Smith, 1956). Collagen shrinks and denatures around 65°C, contributing to the

toughening of meat during cooking; however, if held above 70°C for extended periods,

denatured collagen will begin to gelatinize and increase tenderness (Purslow, 2005, Bailer and

Light, 1989). For this reason, rate of cooking has also shown to play a significant role in the

tenderness of cooked beef. Therefore, the objective of this study was to evaluate the influence of

relative humidity and oven temperature on external and internal color appearance, protein

denaturation, collagen content, shear force values, and sensory attributes of beef strip steaks

cooked using varying oven temperatures and relative humidity levels.

Materials and Methods

Sample Collection, Fabrication, and Treatment Designation

The study was designed as a 2 × 3 factorial utilizing 2 oven temperatures (80°C and 204°C) and 3 levels of relative humidity [zero (ZH), mid (MH), and high (HH)] for a total of 6 individual cooking treatments. In order to maximize humidity level at each oven temperature, different percentages of relative humidity were utilized at each oven temperature because a relative humidity of 100% was unobtainable at 204°C. At 80°C, relative humidity of 0%, 50%, and 100% were utilized for ZH, MH, and HH treatments, respectively. Whereas, at 204°C, relative humidity of 0%, 35%, and 70% were utilized for ZH, MH, and HH treatments, respectively. Therefore, the addition of humidity was relative to the maximal achievable humidity level at each oven temperature. Paired steaks representing each factor combination (randomized within each strip loin) served as the experimental unit. Thirty USDA Low Choice beef strip loins were randomly selected from a commercial beef harvest facility for inclusion in this study. Following collection, all strip loins were transported under refrigeration (2°C) to the Colorado State University Meat Laboratory and stored (2°C) until being fabricated into twelve 2.54 cm steaks. Two adjoining paired steaks (N = 180) were randomly assigned to 1 of 6 treatments methods, so each treatment was represented within each strip loin. The first paired steak was identified as a "shear force" steak, while the second steak was identified as a "sensory" steak. All steaks were vacuum packaged (Clarity Vacuum Pouches #75001839, Koch Supplies, Kansas City, MO), aged for 14 d post mortem at 2°C, then frozen (-20°C) until analysis.

Cooking Procedures

The treatment combinations outlined above were achieved by setting oven temperature (dry bulb temperature) and relative humidity levels (percent moisture) in a commercial

combination oven (Model SCC WE 61 E; Rational, Landsberg am Lech, Germany). Before cooking, frozen steaks were tempered at 2°C for 16-24 h. Steaks were first cross-marked on a 315°C open-hearth char broiler (2 minutes per side), then cooked on a perforated pan at the prescribed oven conditions to an internal temperature of 71°C. Internal steak temperature was monitored throughout the entire oven cooking process using a built in probe placed in the center of a representative steak and final peak temperature was recorded for each steak once removed from the oven (Splash-Proof Super-Fast® Thermapen®, ThermoWorks, Lindon, UT).

External and Internal Steak Appearance and Slice Shear Force Measurements

Steaks designated for shear force were cooked per their respective treatments and subjected to slice shear force (SSF) procedures. Before cooking, each steak was individually weighed to calculate cook loss. Steaks were cooked in batches during 4 individual cooking cycles per treatment. For each cooking cycle, the time required for the entire batch to reach the target temperature was recorded. Upon removal from the oven, each cooked steak was weighed in order to calculate the percent cook loss. Immediately following cooking and before shear force determinations, the external and internal appearance of steaks was evaluated. A colorimeter equipped with a 6 mm measurement port, calibrated at an illuminant of D_{65} and operated at a 10° standard observer angle (Hunter Associates Laboratory, Reston, VA) was used to collect $L^* a^* b^*$ measurements on the exterior and interior of each steak immediately after cooking. Three measurements of $L^* a^* b^*$ were obtained from separate locations within or on the outer surface of the steak to gather an average for each sample. Exterior measurements were taken between char marks created by grill marking the steaks and interior measurements were taken from the most interior portion of the steak cross section at a point 5 cm from the lateral end of the steak. Subjective measurements for degree of doneness, internal, and external steak appearance was

also recorded by 2 trained individuals at the aforementioned locations. Visual degree of doneness was evaluated and recorded using a 5-point scale (1 = rare, 2 = medium rare, 3 = medium, 4 = medium well, and 5 = well done) in reference to published photographic standards (AMSA, 2012). Internal steak appearance was recorded using an 8-point scale (1 = purple, 2 = red, 3 = reddish-pink, 4 = pink, 5 = pinkish-grey, 6 = light brown, 7 = medium brown, and 8 = dark brown). External steak measurements were recorded using an 8-point scale (1 = light grey, 2 = grey, 3 = greyish-brown, 4 = light brown, 5 = brown, 6 = dark brown, 7 = brownish-black, 8 = black).

Slice shear force measurements were obtained from every steak using procedures described by Lorenzen et al. (2010). Within 5 min of recording peak internal temperature, a 1 cm × 5 cm slice was removed from the steak parallel to the muscle fibers from the lateral end and sheared perpendicular to the muscle fibers, using a universal testing machine (Instron Corp., Canton, MA) equipped with a flat, blunt-end blade (crosshead speed: 500 mm/min, load capacity: 100 kg), resulting in a single SSF measurement for each steak. All remaining portions from shear force analysis were saved and frozen (-20°C) for collagen analysis.

Trained Sensory Analysis

Due to limited oven capacity to accommodate 6 cooking treatments, all sensory steaks were cooked in advanced and reheated on the day of analysis. Steaks were cooked following the same cooking protocol as mentioned above. Immediately following cooking, each steak was placed in a vacuum bag, chilled in an ice water bath for 5-15 min, vacuum packaged, and stored at 2-4°C for 16-48 h prior to analysis. To ensure steaks were not becoming excessively oxidized and warmed over during the storage and reheating process, panelists were trained to evaluate oxidized flavor notes to assess during sensory evaluation. On the day of sensory analysis, steaks

were reheated in a circulating water bath set at 57.5°C for 30 minutes. Once removed from the water bath, steaks were trimmed of all external fat and connective tissue, sized into 1 cm cubes, and served to trained panelists. All panelists were trained to evaluated initial tenderness, sustained tenderness, overall tenderness, juiciness, beefy/brothy, browned/grilled, buttery/fat, burnt, bloody/metallic, livery, and oxidized flavors adapted from Adhikari et al. (2011). Each panelist ($n = 7$-8 per session) received 2-3 cubes and evaluated each sample for the aforementioned sensory characteristics using a 10 cm structured line scale verbally anchored at both ends (0 = very tough, very dry, not present; 10 = very tender, very juicy, very intense). Two samples per treatment were served each panel for a total of 12 samples per panel. Two panels were conducted per day: one in the morning and one in the afternoon. Panelists were provided with unsalted saltine crackers, apple juice, and water to cleanse their palate between each sample. All remaining cubes were vacuum packaged and stored at -20°C for protein denaturation analysis.

Protein Denaturation

Denaturation of major skeletal muscle proteins (myosin, sarcoplasmic proteins/collagen, and actin) was evaluated by differential scanning calorimetry (DSC; TA Instruments DSC Q20, Albuquerque, NM). When analyzed via DSC, meat samples produce 3 very distinct denaturation peaks based on denaturation temperature for the aforementioned protein groups, allowing for differentiation in protein groups when analyzing results (Findlay et al., 1986). Denaturation was assessed from remaining cooked sample used for trained sensory analysis. Five strip loins were randomly selected to be analyzed for protein denaturation, with each of the 6 cooking treatments being evaluated per strip loin. An aliquot of 4-10 mg was extracted from the center most portion of each cooked cube and sealed in a DSC pan. An empty pan was used as a reference. The

sample and reference pans were heated from 25°C to 100°C at a heating rate of 5°C/minute. The

peak temperature and denaturation enthalpy (Δ_H) were determined from the DSC curve that was

obtained from each run. Each sample was extracted and analyzed separately in triplicate. The

weight of each sample was used to calculate the change in energy (measured in Joules; J) per g

of sample required to denature remaining intact protein. A greater Δ_H (J/g) is indicative of a

greater amount of intact (undenatured) protein remaining in the sample after the cooking process.

Collagen

Retained samples from the shear force analyses were composited for determination of

total collagen content. Five ($n = 5$), 6 steak composites from each treatment were homogenized

in liquid nitrogen for collagen analysis. Homogenates were prepared and hydroxyproline content

was determined according to the method described by Switzer (1991) using a spectrophotometer.

Collagen content was calculated by multiplying the hydroxyproline content by a factor of 7.52

(Cross et al., 1973).

Statistical Analysis

Data were analyzed using the procedures of SAS (Version 9.4; SAS Inst. Inc., Cary, NC).

The experiment was designed as a 2 × 3 factorial with oven temperature and added humidity as

the fixed effects. Main effect and interaction comparisons were tested for significance using

PROC GLIMMIX with $\alpha = 0.05$ and the denominator degree of freedom was calculated by the

Kenward-Roger method. For trained sensory analysis, the scores of each panelist were averaged,

resulting in one value per sample and panel number was included in the model as a random

variable. Peak cooking temperature of each steak was initially included as a covariate in each

model but was removed from final models due to a lack of significance.

Results and Discussion

Cooking Rate and Cook Loss

The time required to cook beef strip steaks to 71°C for each of the 6 treatments is presented in Table 1. At 80°C, MH and HH decreased cooking time by 93.25 and 107.75 min, respectively, compared to steaks cooked with ZH. Cooking times were shorter when relative humidity was added at 204°C; however, it was to a much lower extent. At 204°C, MH and HH decreased cooking time by 3.75 and 7.00 min, respectively, compared to steaks cooked with ZH. Evaluating various beef roasts, Jeremiah and Gibson (2003) produced similar results showing very noticeable decreases in cooking time when added moisture was included in the cooking process. Water has a much greater specific heat capacity than air, thus, adding moisture to the cooking environment significantly increased the efficiency of heat transfer from the environment to the steak to decrease cooking time. These results agree with previous studies reporting an increase in cooking rate when humidity is added to the cooking environment (Laakkonen et al., 1970; Vittadini et al., 2005). Additionally, cooking times for 80°C-HH and 204°C-ZH steaks were essentially equal (17.53 vs 17.00 min, respectively), which allowed for the inadvertent comparison of the two treatments to assess the influence of humidity when cooking rate is kept similar. Percent cook loss was influenced by an oven temperature $\times$ relative humidity interaction (Table 2). Steaks cooked at 80°C-ZH showed the greatest ($P < 0.01$) loss of moisture during cooking of all treatments. Relative humidity had no ($P > 0.05$) impact on cook loss when steaks were cooked at 204°C; nevertheless, cook loss at 204°C was still greater ($P < 0.01$) than both 80°C-MH and 80°C-HH. These results agree with those found by Belk et al. (1993) who showed that increasing oven temperature, adding humidity, and ultimately increasing cooking rate resulted in a decrease in cooking yields. King et al. (2003) concluded rapid cooking rates cause

107

excessive myofibrillar shortening, resulting in decreased cooking yields. Therefore, increased cooking yield by addition of humidity may only be achievable if a slower cooking rate is maintained by a lower oven temperature.

Cooked Steak Color

Instrumental and visual assessment of external and internal color of cooked strip steaks are presented in Table 2. All color measurements were affected by an interaction of the main effects. For both oven temperatures, external $L*$ values increased ($P < 0.01$) with increasing relative humidity levels, which is indicative of a lighter surface color. Additionally, 80°C-ZH steaks had the darkest external color ($P < 0.01$) of all treatments as determined by $L*$ values. When steaks were cooked at 204°C, $a*$ values, which are indicative of a redder color, decreased ($P < 0.01$) as relative humidity increased. However, when cooked at 80°C, $a*$ values were the lowest ($P < 0.01$) when humidity was absent from the cooking environment. Similarly, Isleroglu et al. (2014) reported higher $L*$ and lower $a*$ values from chicken breasts cooked using a steam-assisted hybrid oven. Visual assessment of external color produced similar results as instrumental color values. Trained panelists found 80°C-ZH steaks had the darkest external color ($P < 0.01$) of all treatments. At 80°C, external surface color became lighter ($P < 0.01$) as relative humidity increased. Similarly, external color of 204°C-ZH steaks were darker ($P < 0.01$) than both 204°C-MH and 204°C-HH steaks; however, no visual color differences ($P > 0.05$) were found between 204°C-MH and 204°C-HH samples. These results indicate that added moisture inhibited surface browning during cooking.

Internal $L*$ values were greatest ($P < 0.01$) from 80°C-ZH steaks compared to all other treatments. No additional differences ($P > 0.05$) in internal $L*$ values were observed among any other treatments. Steaks cooked at 80°C-ZH produced the lowest ($P < 0.01$) $a*$ values of all

treatments. Based on instrumental color values, 80°C-ZH steaks had the appearance of being the most well done, regardless of all treatments being cooked to the same internal temperature (71°C). This is also reflected in trained panelist ratings for doneness and internal color. Steaks cooked at 80°C-ZH appeared to be the most ($P < 0.01$) well done, as well as, the brownest ($P < 0.01$) internally of all treatments. No other treatment ($P > 0.05$) differed in the visual assessment of doneness and appeared to be cooked to a medium degree of doneness. Assessing subjective and objective measurements for both external and internal cooked surfaces, it is evident that cooking parameters had a significant influence on the appearance of cooked steaks.

Protein Denaturation

Table 3 shows the change in enthalpy (Δ_H; J/g) and peak temperatures required to denature remaining intact myosin, sarcoplasmic proteins and collagen, and actin. Previously published literature has determined that denaturation temperatures for myosin, sarcoplasmic proteins and collagen, and actin are 55.5°C, 66.8°C, and 80.9°C, respectively (Findlay et al., 1986). In the current study, similar denaturation temperatures were recorded for myosin (56.6°C), sarcoplasmic proteins and collagen (65.4°C), and actin (80.5°C). Because denaturation samples were extracted from the inner-most portion of steak cross-sections and each steak was cooked to the sample end-point temperature, it is speculated that an increase in protein denaturation would indicate a slower transfer of heat from the surface to the interior, facilitating prolonged exposure to denaturation temperatures.

Differences in Δ_H for each protein appear to be related to cooking rate and exposure time to heat, which was influenced by relative humidity level. No differences ($P = 0.86$) in Δ_H were observed for myosin, regardless of relative humidity. The low Δ_H for myosin indicates that nearly all myosin was denatured during the initial cooking process and would not be expected to

contribute to tenderness differences. Myosin shrinks upon denaturation and is responsible for the initial toughening phase of meat during cooking (McCormick, 1999). Most sarcoplasmic proteins and collagen were denatured during cooking with ZH and MH; however, greater ($P = 0.02$) amounts remained intact in steaks cooked with HH. Thus, it seems that the reduction in cooking time of HH steaks did not allow for adequate exposure of sarcoplasmic proteins and collagen to the required denaturation temperature to completely alter the protein structure.

Unlike myofibrillar proteins, sarcoplasmic proteins expand rather than shrink upon heating, as well as, forming aggregates and gelatinizing during cooking (Baldwin, 2012). Furthermore, many of these proteins are enzymes, which have shown to have a tenderizing effect when cooked for long periods of time at low temperatures (Tornberg, 2005). Although denatured enzymes would lose functionality, the current data suggest a slower transfer of heat and the potential for an increased window of opportunity for enzyme activity at lower temperatures, particularly in the more interior portions of the steak.

Collagen begins to denature and shrink between 64-68°C, which results in toughening of meat (McCormick, 1999); however, as meat continues to be held at temperatures greater than 70°C, collagen begins to solubilize and an increase in tenderness is observed (Bailey and Light, 1989). The Δ_H of sarcoplasmic proteins and collagen follows trained sensory tenderness ratings. Lower Δ_H was determined for sarcoplasmic proteins and collagen at ZH and MH, while SSF and sensory tenderness ratings were more favorable at these same humidity levels. Therefore, the differences in the effects of heat on sarcoplasmic proteins and collagen may have played a role in the tenderization of these treatments. Additionally, greater ($P = 0.02$) amounts of actin remained intact in steaks cooked using MH and HH when compared to steaks cooked using ZH; however,

based off Δ_H values, substantial amounts of actin remained intact in all relative humidity levels after the cooking process.

The substantial increase in cooking time of steaks cooked with ZH facilitated an increased exposure of proteins to the higher temperatures required to denature actin. Previously, Bertram et al. (2006) showed that Δ_H of all 3 proteins gradually decreased as final internal temperature increased until peaks were practically devoid when samples were cooked to 75°C. Although the current study did not evaluate different internal temperatures, cooking procedures from both studies facilitated differences in heat transfer which appeared to have comparable effects on protein denaturation. Differential scanning calorimetry has been widely used to evaluate thermal properties of meat proteins; however, a definitive relationship between DSC thermograms and meat tenderness has yet to be established.

Collagen

Concentrations of total collagen content remaining in cooked steaks are shown in Table 4. It has been suggested that any collagen gelatinized during cooking would be released with the cook loss (Palka, 1999), thus it was assumed that any changes in collagen solubility would be observed by measuring total collagen remaining in cooked steaks. Steaks cooked with ZH, regardless of oven temperature, had greater ($P < 0.01$) concentrations of collagen than MH and HH steaks. Total collagen content did not have much of an effect on sensory tenderness, since ZH steaks had the greatest concentrations of collagen, but were some of the most tender steaks. Also, the greater collagen content of ZH steaks may have been partially related to the greater cook loss seen in 80°C-ZH steaks, resulting in a more concentrated collagen content. Although many studies have found significant relationships between total collagen and tenderness (Riley et al., 2005), others have been inconsistent in attempting to fully understand and establish the

relationship between connective tissue and meat tenderness (Reagan et al., 1976; Seideman et al., 1987). Furthermore, tenderness of strip steaks is affected less by collagen content and more by proteolysis (Koohmaraie and Geesink, 2006), so any differences in total collagen content may have had a smaller influence on perceived tenderness. Although collagen content was greatest in ZH treatments, this may not have reflected the heat-induced structural changes in collagen that may have occurred due to an extended cooking time and exposure to temperatures greater than 70°C. Differences in sensory tenderness due to oven temperature were not explained by protein denaturation results alone, suggesting that changes in the structure of collagen may have very well contributed to differences in tenderness. Consequently, there could be benefit in using techniques to more specifically evaluate the heat-induced structural changes in collagen due to varying oven temperatures, relative humidity levels, and cooking rates in future studies.

Shear Force

Slice shear force values were affected by relative humidity (Table 4). Regardless of oven temperature, steaks cooked with HH produced greater ($P = 0.02$) SSF values than steaks cooked with both ZH and MH. Berry et al. (1977) did not find shear force differences between oven roasted and braised beef *semimembranosus* but did observe lower sensory tenderness ratings for braised samples. Tenderness decreases as cooking rate increases (Cross et al., 1976). In the present study, the increase in SSF values are believed to be the result of an increased cooking rate by the addition of high levels of humidity in the cooking environment. However, the utilization of moderate levels of humidity facilitated a more rapid cooking rate without negatively affecting shear force values.

Trained Sensory

Tenderness ratings were influenced by both oven temperature and relative humidity (Table 5). Trained panelists rated steaks cooked at 80°C greater ($P < 0.01$) than those cooked at 204°C for initial tenderness, sustained tenderness, and overall tenderness. Previous literature has shown that reducing oven temperature and cooking rate results in a more tender beef product (King et al., 2003; Christensen et al., 2011). Additionally, steaks cooked with ZH and MH were rated greater ($P < 0.01$) than those cooked with HH for initial tenderness, sustained tenderness, and overall tenderness. Again, decreased tenderness appeared to be related to cooking rate. Unlike the current findings, many previous studies have reported increases in tenderness as a result of moist-heat cookery when compared to dry-heat cookery; however, many of these studies used moist-heat methods with a slower cooking rate (Kolle et al., 2004) than dry-heat methods used. Therefore, it can be difficult to determine if differences were due to added moisture or cooking rate. In agreement with the current results, however, Berry et al. (1977) found semimembranosus steaks cooked using dry-heat to be more tender than those cooked with moist-heat.

Trained sensory ratings for juiciness were affected by an interaction of the main effects (Table 5). When steaks were cooked at 204°C, relative humidity had no influence ($P > 0.05$) on juiciness. However, when steaks were cooked at 80°C, steaks cooked with MH were rated juicier ($P < 0.01$) than both HH and ZH, respectively. Furthermore, 80°C-ZH steaks were rated as the least juicy ($P < 0.01$) of all treatments and 80°C-MH steaks the juiciest ($P < 0.01$). As expected, juiciness scores followed cook loss percentages. Previous studies have shown that moist-heat cookery yields a juicier product due to a decrease in evaporative moisture loss (Bowers et al., 2012).

Beefy/brothy, buttery/fat, and bloody/metallic flavor intensities were influenced by an oven temperature × relative humidity interaction (Table 5). When cooked at 204°C, relative humidity had no influence ($P > 0.05$) on ratings for beefy/brothy intensity; but, when steaks were cooked at 80°C, ZH produced the most intense ($P < 0.01$) beefy/brothy flavors. Furthermore, 80°C-ZH samples had the greatest beefy/brothy intensity ($P < 0.01$) of all treatments. Cooking steaks with ZH also produced the least intense ($P = 0.04$) bloody/metallic flavors, regardless of oven temperature. Adding relative humidity increased ($P = 0.04$) bloody/metallic intensity for both oven temperatures; however, 80°C-MH and 80°C-HH steaks had a more intense ($P = 0.04$) bloody/metallic flavor than both 204°C-MH and 204°C-HH steaks. Cooking with HH decreased ($P = 0.01$) buttery/fat intensity at 204°C; but, similar trends were not seen at 80°C. Steaks cooked at 80°C-MH produced a more intense ($P = 0.01$) buttery/fat flavor than 80°C-ZH steaks, with 80°C-MH steaks performing similarly ($P > 0.05$) to both relative humidity levels.

Both browned/grilled and burnt flavor intensities were affected ($P < 0.01$) by relative humidity. Zero humidity steaks produced the most ($P < 0.01$) and HH steaks the least ($P < 0.01$) intense ratings for browned/grilled flavors. Browned/grilled flavors are associated with Maillard reaction products that occur on the surface of cooked meat (Mottram, 1998) and these reactions are inhibited in moist cooking environments (Kerth and Miller, 2015). It is evident in the current study, by evaluating external color and sensory scores, that increasing moisture to the cooking environment inhibited the non-enzymatic browning process. This conclusion is further supported by Isleroglu, et al. (2014) who measured a decrease in the production of Maillard reaction products after adding humidity to the cooking environment of chicken.

No differences ($P > 0.05$) were observed among burnt, livery, or oxidized off-flavors (Table 5). Ratings for livery and oxidized intensity were low in some samples but were not

114

present in most samples evaluated by panelists. These low intensity ratings would not be

expected to play a role in the flavor perception of consumers. In the current study, intensity

ratings for livery and oxidized were collected to ensure that these off-flavors were not introduced

during the reheating of steaks in the water bath. Oxidation can occur during the storage and

reheating of cooked meat, which is commonly referred to as "warmed over flavor" (Mottram,

1998); however, vacuum packaging steaks immediately after cooking and reheating in the

absence of oxygen prevented oxidation and the development of warmed over flavor in sensory

steaks.

Although it was not the objective of the treatment design, 80°C-HH (100% humidity) and

204°C-ZH (dry-heat) treatments had comparable cooking rates (17.53 versus 17.00 min,

respectively). This allowed for the evaluation of the effects of adding 100% humidity to the

cooking environment on sensory development when cooking rates were similar. Neither initial,

sustained, nor overall tenderness differed between dry-heat and 100% humidity cooking;

however, cooking with 100% humidity produced juicier ($P < 0.01$) steaks. Greater flavor

intensities were recorded for dry-heat cooking, as panelists rated steaks cooked using dry-heat

more ($P < 0.01$) intense for beefy/brothy, browned/grilled, and buttery/fat flavors; and less ($P <$

0.01) for bloody/metallic. Although juiciness was more favorable for 100% humidity cooking,

dry-heat cooking produced more favorable flavors. While consumer sensory and acceptability

was not evaluated in the current study, recent consumer studies have found flavor to be the most

influential eating quality trait when determining consumer overall acceptability of beef (O'Quinn

et al., 2012; Hunt et al., 2014; Legako et al., 2016), particularly when tenderness is acceptable.

Therefore, it is suggested that the advantages in the flavor of dry-heat cooked steaks would be

more desirable to consumers over steaks cooked in the presence of 100% humidity, when

cooking rate is kept similar. Based on these observations, future work emphasizing the influence of added humidity on sensory development would provide further insight into possible factors affecting these variables.

Conclusions

Adding humidity to the cooking environment was expected to improve tenderness of beef strip steaks; however, the current findings show humidity had a negative effect on tenderness development. It appears that tenderness was affected more by cooking rate, which was altered by both oven temperature or humidity level. Adding humidity increased the cooking rate of strip steaks at both oven temperatures, but this difference was substantially more evident at 80°C. A slower cooking rate is believed to be responsible for a more tender product because of an increased exposure time to heat, better facilitating the thermal breakdown of proteins and likely the solubilization of collagen. Adding moderate levels of humidity to the cooking environment improved the efficiency of the cooking process without affecting tenderness attributes. Cooking at 80°C with no humidity produced a tender product; however, it was exceptionally dry and produced more roast-like flavors that may not be desirable when consuming a steak product. At 80°C, the addition of 50% humidity allowed for a drastic decrease in cooking time without sacrificing tenderness and juiciness; however, it hindered the development of browned/grilled flavors. At 204°C, the addition of 35% humidity decreased cooking time while only minimally affecting browned/grilled flavor development. Further work is warranted to understand how the observed differences in tenderness and flavor attributes would influence consumer acceptability.

Table 5. 1 Interaction means for the length of time required to cook beef strip steaks to 71°C and cook loss using two oven temperatures and three levels of humidity.

	Cook Time (min)	Cook Loss (%)
80°C		
Zero Humidity	125.29[a]	33.03[a]
Mid Humidity	32.03[b]	21.37[c]
High Humidity	17.53[c]	21.91[c]
204°C		
Zero Humidity	17.00[c]	24.56[b]
Mid Humidity	13.53[c]	24.41[b]
High Humidity	10.60[c]	24.64[b]
SEM[1]	2.48	0.71
P – Value	< 0.01	< 0.01

[a-c]Means in the same column lacking a common superscript differ ($P < 0.05$)
[1]Standard error (largest) of the least squares mean

Table 5. 2 External and internal color (CIE $L*$, $a*$, and $b*$) and trained personnel (n = 2) visual assessment of doneness, external color, and internal color of beef strip steaks cooked to 71°C using two oven temperatures and three levels of humidity.

Color Measurement	80°C			204°C				P – Value		
	Zero Humidity	Mid Humidity	High Humidity	Zero Humidity	Mid Humidity	High Humidity	SEM[1]	Oven Temp	Relative Humidity	OT × RH
External Color										
$L*$	24.30[d]	36.83[b]	41.90[a]	33.13[c]	37.30[b]	41.66[a]	0.85	< 0.01	< 0.01	< 0.01
$a*$	9.38[e]	13.04[cd]	11.85[d]	16.89[a]	15.14[b]	13.18[c]	0.47	< 0.01	< 0.01	< 0.01
$b*$	9.81[c]	18.96[b]	18.18[b]	24.12[a]	22.79[a]	20.19[b]	0.80	0.43	< 0.01	< 0.01
Internal Color										
$L*$	54.69[a]	49.01[b]	49.37[b]	48.52[b]	49.08[b]	48.86[b]	0.64	< 0.01	< 0.01	< 0.01
$a*$	11.80[c]	18.65[ab]	16.18[b]	18.66[ab]	17.36[ab]	19.77[a]	1.31	< 0.01	0.04	< 0.01
$b*$	16.59[c]	18.89[a]	17.14[bc]	18.73[a]	17.74[b]	16.67[c]	0.34	0.54	< 0.01	< 0.01
Visual Assessment										
Doneness[2]	4.81[a]	3.26[b]	3.33[b]	3.27[b]	3.18[b]	3.28[b]	0.09	< 0.01	< 0.01	< 0.01
External Color[3]	7.13[a]	5.11[bc]	4.68[d]	5.41[b]	4.90[cd]	4.60[d]	0.11	< 0.01	< 0.01	< 0.01
Internal Color[4]	6.78[a]	5.05[c]	5.50[b]	5.27[bc]	5.46[b]	5.45[b]	0.09	< 0.01	< 0.01	< 0.01

[a-d] Means in the same row lacking a common superscript differ ($P < 0.05$)
[1] Standard error (largest) of the least squares means
[2] Doneness: 1 = rare, 2 = medium rare, 3 = medium, 4 = medium well, 5 = well done
[3] External Color: 1 = light grey, 2 = grey, 3 = greyish-brown, 4 = light brown, 5 = brown, 6 = dark brown, 7 = brownish-black, 8 = black
[4] Internal Color: 1 = purple, 2 = red, 3 = reddish-pink, 4 = pink, 5 = pinkish-grey, 6 = light brown, 7 = medium brown, and 8 = dark brown

Table 5. 3 Change in enthalpy[1] required to denature remaining intact myosin, sarcoplasmic protein, and actin of beef strip steaks cooked to 71°C using three levels of humidity.

Relative Humidity Level	Myosin (J/g)	Sarcoplasmic Protein and Collagen (J/g)	Actin (J/g)
Zero Humidity	0.047	0.034[b]	0.307[a]
Mid Humidity	0.034	0.035[b]	0.734[b]
High Humidity	0.027	0.121[a]	0.646[b]
Peak Denaturation Temperature (°C)	56.661	65.410	80.518
SEM[2]	0.038	0.024	0.161
P - Value	0.866	0.020	0.023

[a-b]Means in the same column lacking a common superscript differ ($P < 0.05$)
[1]Change in enthalpy presented as change in Joules (J) per g required to denature remaining intact protein
[2]Standard error (largest) of the least squares means

Table 5. 4 Slice shear force (SSF) values, total collagen content (dry matter basis), and trained sensory ratings[1] for overall tenderness of beef strip steaks cooked to 71°C using two oven temperatures and three levels of humidity.

	SSF (kg)	Collagen (mg/g)	Overall Tenderness
Oven Temperature			
80°C	16.00	15.98	6.50^a
204°C	16.78	13.65	6.09^b
SEM[1]	0.53	0.92	0.08
P - Value	0.30	0.08	< 0.01
Added Humidity			
Zero Humidity	15.70^m	19.36^m	6.47^m
Mid Humidity	15.59^m	12.66^n	6.42^m
High Humidity	17.88^n	12.44^n	6.01^n
SEM[2]	0.65	1.59	0.14
P - Value	0.02	< 0.01	< 0.01
OT × RH P – Value	0.85	0.18	0.81

[a-b]Means in the same column lacking a common superscript differ ($P < 0.05$) due to oven temperature

[m-n]Means in the same column lacking a common superscript differ ($P < 0.05$) due to added humidity

[1]Attributes were scored using a 10 cm structured line scale: 0 = very tough; 10 = very tender

[2]Standard error (largest) of the least squares means

Table 5. 5 Trained sensory ratings[1] for beef strip steaks cooked to 71°C using two oven temperatures and three levels of humidity.

	80°C			204°C				*P* – Value		
Trait	Zero Humidity	Mid Humidity	High Humidity	Zero Humidity	Mid Humidity	High Humidity	SEM[2]	Oven Temp	Added Humidity	OT ✕ RH
Initial Tenderness	6.86^{am}	6.82^{am}	6.41^{an}	6.39^{bm}	6.47^{bm}	6.14^{bn}	0.13	< 0.01	< 0.01	0.70
Sustained Tenderness	6.53^{am}	6.38^{am}	5.92^{an}	5.56^{bm}	6.00^{bm}	5.56^{bn}	0.14	< 0.01	< 0.01	0.73
Overall Tenderness	6.72^{am}	6.62^{am}	6.17^{an}	6.21^{bm}	6.22^{bm}	5.84^{bn}	0.14	< 0.01	< 0.01	0.81
Juiciness	4.18^{z}	6.10^{w}	5.72^{x}	5.36^{y}	5.28^{y}	5.23^{y}	0.12	0.63	< 0.01	< 0.01
Beefy/Brothy	5.63^{w}	5.07^{xy}	4.97^{y}	5.26^{x}	5.15^{xy}	5.08^{xy}	0.15	0.38	< 0.01	0.01
Browned/Grilled	4.79^{m}	4.24^{n}	4.10^{o}	4.73^{m}	4.61^{n}	4.28^{o}	0.20	0.09	< 0.01	0.19
Buttery/Fat	2.22^{y}	2.56^{wx}	2.33^{xy}	2.74^{w}	2.63^{w}	2.23^{y}	0.11	0.05	0.01	0.01
Burnt	0.65	0.50	0.64	0.66	0.39	0.67	0.14	0.84	0.15	0.82
Bloody/Metallic	1.10^{y}	1.90^{w}	1.66^{wx}	1.09^{y}	1.34^{y}	1.37^{xy}	0.11	< 0.01	< 0.01	0.04
Livery	0.07	0.02	0.03	0.01	0.02	0.07	0.01	0.73	0.32	0.08
Oxidized	0.06	0.05	0.02	0.05	0.11	0.11	0.03	0.10	0.80	0.41

[a-b]Least square means in the same row without a common superscript differ (*P* < 0.05) due to oven temperature

[m-o]Least square means in the same row without a common superscript differ (*P* < 0.05) due to added humidity

[w-z]Least square means in the same row without a common superscript differ (*P* < 0.05) due to an over temperature ✕ realtive humidity interaction

[1]Attributes were scored using a 10 cm structured line scale: 0 = very tough, very dry, and not present; 10 = very tender, very juicy, and very intense

[2]Standard error of the least squares means

LITERATURE CITED

Adhikari, K., E. Chambers IV, R. Miller, L. Vázquez-Araújo, N. Bhumiratana, and C. Philip. 2011. Development of a lexicon for beef flavor in intact muscle. J. Sens. Stud. 26:413–420.

American Meat Science Association (AMSA). 2012. Meat Color Measurement Guidelines. Am. Meat Sci. Assoc., Champaign, IL.

Baldwin, D. E. 2012. Sous vide cooking: A review. Int. J. Gastron. Food Sci. 1:15–30.
Berry, B. W., M. R. Wheeling, and J. A. Carpenter. 1977. Effects of muscle and cookery method on palatability of beef from several breeds and breed crosses. J. Food Sci. 42:1322–1324.

Bertram, H. C., Z. Wu, F. van den Berg, and H. J. Andersen. 2006. NMR relaxometry and differential scanning calorimetry during meat cooking. Meat Sci. 74:684–9.

Berry, B. W., M. R. Wheeling, and J. A. Carpenter. 1977. Effects of muscle and cookery method on palatability of beef from several breeds and breed crosses. J. Food Sci. 42:1322–1324.

Bowers, L. J., M. E. Dikeman, L. Murray, and S. L. Stroda. 2012. Cooked yields, color, tenderness, and sensory traits of beef roasts cooked in an oven with steam generation versus a commercial convection oven to different endpoint temperatures. Meat Sci. 92:97–106.

Christensen, L., P. Ertbjerg, M. D. Aaslyng, and M. Christensen. 2011. Effect of prolonged heat treatment from 48 °C to 63 °C on toughness, cooking loss and color of pork. Meat Sci. 88:280–5.

Cross, H. R., Z. L. Carpenter, and G. C. Smith. 1973. Effect of Intramuscular Collagen and Elastin on Bovine Muscle Tenderness. J. Food Sci. 38:998–1003.

Cover, S., and W. H. Smith. 1956. The effect of two methods of cooking on palatability scores, shear force values, and collagen content of two cuts of beef. J. Food Sci. 21:312–321.

Findlay, C. J., D. W. Stanley, and E. A. Gullett. 1986. Thermomechanical properties of beef muscle. Meat Sci. 16:57–70.

Hunt, M. R., A. J. Garmyn, T. G. O'Quinn, C. H. Corbin, J. F. Legako, R. J. Rathmann, J. C. Brooks, and M. F. Miller. 2014. Consumer assessment of beef palatability from four beef muscles from USDA Choice and Select graded carcasses. Meat Sci. 98:1–8.

Isleroglu, H., T. Kemerli, Ö. Özdestan, A. Uren, and F. Kaymak-Ertekin. 2014. Effect of oven cooking method on formation of heterocyclic amines and quality characteristics of chicken patties: steam-assisted hybrid oven versus convection ovens. Poult. Sci. 93:2296–303.

Jeremiah, L. E., and L. L. Gibson. 2003. Cooking influences on the palatability of roasts from the beef hip. Food Res. Int. 36:1–9.

Kerth, C. R., and R. K. Miller. 2015. Beef flavor: A review from chemistry to consumer. J. Sci. Food Agr. 95:2783–2798.

King, D. A., M. E. Dikeman, T. L. Wheeler, C. L. Kastner, M. Koohmaraie, D. a. King, M. E. Dikeman, T. L. Wheeler, C. L. Kastner, and M. Koohmaraie. 2003. Chilling and cooking rate effects on some myofibrillar determinants of tenderness of beef. J. Anim. Sci. 81:1473–1481.

Kolle, B. K., D. R. McKenna, and J. W. Savell. 2004. Methods to increase tenderness of individual muscles from beef rounds when cooked with dry or moist heat. Meat Sci. 68:145–54. Koohmaraie, M., and G. H. Geesink. 2006. Contribution of postmortem muscle biochemistry to the delivery of consistent meat quality with particular focus on the calpain system. Meat Sci. 74:34–43.

Laakkonen, E., G. H. Wellington, and J. N. Sherbon. 1970. Low-Temperature, Long-Time Heating of Bovine Muscle 1. Changes in Tenderness, Water-Binding Capacity, pH and Amount of Water-Soluble Components. J. Food Sci. 35:175–177.

Legako, J. F., T. T. N. Dinh, M. F. Miller, K. Adhikari, and J. C. Brooks. 2016. Consumer Palatability Scores, Sensory Descriptive Attributes, and Volatile Compounds of Grilled Beef Steaks from Three USDA Quality Grades. Meat Sci. 112:77–85.

Lorenzen, C. L., C. R. Calkins, M. D. Green, R. K. Miller, J. B. Morgan, and B. E. Wasser. 2010. Efficacy of performing Warner–Bratzler and slice shear force on the same beef steak following rapid cooking. Meat Sci. 85:792-794.

McCormick, R. J. 1999. Extracellular modifications to muscle collagen: implications for meat quality. Poultry Sci. 78:785–791.

Mottram, D. S. 1998. Flavour formation in meat and meat products: a review. Food Chem. 62:415–424.

O'Quinn, T. G., J. C. Brooks, R. J. Polkinghorne, A. J. Garmyn, B. J. Johnson, J. D. Starkey, R. J. Rathmann, and M. F. Miller. 2012. Consumer assessment of beef strip loin steaks of varying fat levels. J. Anim. Sci. 90:626–634.

Palka, K. 1999. Changes in intramuscular connective tissue and collagen solubility of bovine m. semitendinosus during retorting. Meat Sci. 53:189–194.

Purslow, P. P. 2005. Intramuscular connective tissue and its role in meat quality. Meat Sci 70:435–447.

Reagan, J. O., Z. L. Carpenter, and G. C. Smith. 1976. Age-Related Traits Affecting the Tenderness of the Bovine Longissimus Muscle. J. Anim. Sci. 43:1198–1205.

Riley, D. G., D. D. Johnson, C. C. Chase, R. L. West, S. W. Coleman, T. A. Olson, and A. C. Hammond. 2005. Factors influencing tenderness in steaks from Brahman cattle. Meat Sci. 70:347–56.

Seideman, S. C., M. Koohmaraie, and J. D. Crouse. 1987. Factors associated with tenderness in young beef. Meat Sci. 20:281–91.

Switzer, B. 1991. Determination of hydroxyproline in tissue. J. Nutr. Biochem. 2:229–231.

Tornberg, E. 2005. Effects of heat on meat proteins - Implications on structure and quality of meat products. Meat Sci. 70:493–508.
Vittadini, E., M. Rinaldi, E. Chiavaro, D. Barbanti, and R. Massini. 2005. The effect of different convection cooking methods on the instrumental quality and yield of pork Longissimus dorsi. Meat Sci. 69:749–756.

Yancey, J. W. S., M. D. Wharton, and J. K. Apple. 2011. Cookery method and end-point temperature can affect the Warner-Bratzler shear force, cooking loss, and internal cooked color of beef longissimus steaks. Meat Sci. 88:1–7.