



## Final Report of the College of Engineering Investigation Committee Concerning Allegations of Research Misconduct

July 31, 2018

### Introduction

A College of Engineering Investigation Committee (COEIC) was formed under the University Policy and Procedures Concerning Research Misconduct<sup>1</sup> (the “Policy”) on September 18, 2017 to address allegations of possible Research Misconduct made against current University Professor, Dr. Mingjun Zhang, Department of Biomedical Engineering. An initial allegation of possible misrepresentation (“falsification”) of the identity of the newly identified gene for an arabinogalactan protein from English Ivy (IAGP) in a manuscript published in 2016 was made by an anonymous Complainant and submitted to the University on June 23, 2016.

Allegation #1 – The identity and species of origin of the Ivy arabinogalactan protein (IAGP) was falsified by Dr. Zhang by the intentional, knowing, and/or reckless misrepresentation of data, text, and figures related to the isolation and characterization of the clone identified, published in Fig. 3 and Fig. S4A of PNAS 113(23):E3193-3202 and deposited in GenBank database as (KM820289).

Based on recommendations from the Committee of Initial Inquiry (CII), a second allegation was added and included in the allegations to be investigated by the COEIC.

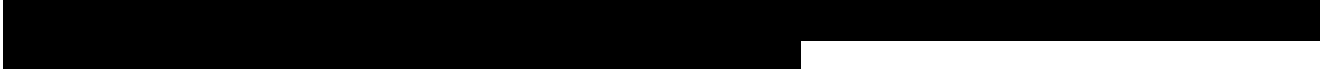
Allegation #2 – MALDI-TOF mass spectrometry data was falsified by Dr. Zhang by the intentional, knowing, and/or reckless misrepresentation of the published methods used including the location where the MALDI-TOF mass spectrometry experiments were conducted in the text and Fig. S4B of PNAS 113(23):E3193-3202.

The allegations referred to figures and text included in the below publication:

Yujian Huang, Yongzhong Wang, Li Tan, Leming Sun, Jennifer Petrosino, Mei-Zhen Cui, Feng Hao and Mingjun Zhang (2016). Nanospherical arabinogalactan proteins are a key component of the high-strength adhesive secreted by English Ivy. PNAS 113 (23): E3193-3202. (herein after referred to as “PNAS 2016”<sup>2</sup>).

[Cited funding: Army Research Office (ARO) W911NF-10-1-0114 and ARO W911NF-12-1-0294; National Science Foundation (NSF) Civil, Mechanical, and Manufacturing Innovation #1029953, NSF Chemical Bioengineering, Environmental and Transport Systems # 0965877; Department of Energy (DOE) Bioenergy Science Center #DE-AC05-00OR22725]

### Other Respondents regarding these allegations:

- 

<sup>1</sup> ATT-1 - University Policy and Procedures Concerning Research Misconduct

<sup>2</sup> ATT-2 - PNAS 113 23 E3193-3202 + SI



## **Preliminary Assessment Summary**

On July 12, 2016, Dr. Yucel, Dr. Richard Hart, Chair, department of Biomedical Engineering, and Dr. Randy Moses, Associate Dean for Research, met to review the allegations and to conduct a Preliminary Assessment<sup>3</sup> as required by the Policy. Given the specific domain expertise represented in the allegations, it was determined that a subject matter expert should be consulted to assist in the determination as to whether the allegations might indicate possible Research Misconduct versus simply a difference of scientific opinion or the interpretation of the research results.

On July 13, 2016, Dr. Mike Ibba, Professor and Chair, department of Microbiology, College of Arts and Sciences, was enlisted to act as a consultant on this matter. Dr. Ibba was asked to review the allegations and provide his professional opinion as to whether the allegations may indicate possible Research Misconduct or if they represented something else, for instance a difference of scientific opinion or the interpretation of research results.

On July 15, 2016, Dr. Etta Kavanagh, Editorial Manager for PNAS, contacted the University regarding a complaint received by the journal. Dr. Yucel determined that the complaint sent to the journal was the same as the complaint received by the University. Dr. Yucel contacted Dr. Kavanagh on July 21, 2016 acknowledging the University's receipt of the complaint and verifying that the university was reviewing the matter. Dr. Yucel also requested that the journal provide any correspondence between the journal and Dr. Zhang related to the manuscript in question. On July 25, 2016, Dr. Kavanagh provided copies of four decision letters sent by PNAS to the authors corresponding to the multiple submissions to that journal. Dr. Yucel then sent an additional request to Dr. Kavanagh for copies of any submissions by the authors in response to reviewer's comments. These were provided to Dr. Yucel on July 26, 2016 by the journal.

On July 26, 2016, Drs. Yucel, Hart and Moses met to review the information provided by the journal, Dr. Ibba's assessment, and some basic sequence and homology analyses conducted by Dr. Yucel. Based on this review, it was determined that the allegation of possible Research Misconduct (Falsification) was credible and specific enough to indicate possible Research Misconduct and should be moved forward to an Inquiry.

The Preliminary Assessment letter regarding the allegations of potential Research Misconduct and recommending that this case be moved forward to a Committee of Initial Inquiry was submitted to Dr. Caroline Whitacre, Senior Vice President for Research on August 25, 2016. On September 6, 2016, Dr. Whitacre concurred with the Preliminary Assessment and indicated a Committee of Initial Inquiry should be initiated.

## **Notification of Allegations to Respondent(s)**

On August 23, 2016, Dr. Yucel, Dr. Moses, and Dr. Hart met with Dr. Mingjun Zhang, senior and corresponding author of the PNAS 2016 paper, to notify him of the allegations made against the paper, to discuss the allegations and determine who had participated in the generation of the data in question. This meeting was recorded by Dr. Yucel and transcribed.<sup>4</sup> During this meeting, Dr. Zhang indicated that [REDACTED]

[REDACTED] Based on the information provided by Dr. Zhang during this meeting, it was determined that the allegations indicating possible Research Misconduct should be expanded to include [REDACTED] as a Respondent. Immediately following the discussion with Dr. Zhang, Dr. Yucel met with [REDACTED] to notify [REDACTED] of the allegations and to let [REDACTED] know

<sup>3</sup> ATT-3 - 20160825 – Preliminary Assessment – Zhang [REDACTED]

<sup>4</sup> ATT-4 - 20160823 – Transcript Notification of Allegations Zhang #1



that [REDACTED] would also be considered a Respondent in the case. This meeting was recorded by Dr. Yucel and transcribed.<sup>5</sup>

### Sequestration of Data

Immediately following the meetings on August 23, 2016, all relevant research records were sequestered from Dr. Zhang's laboratory. Forensic images for three identified computer systems were obtained over August 23 - 24, 2016. [REDACTED]

### Committee of Initial Inquiry (CII) Summary

A Committee of Initial Inquiry ("CII") was formed on September 6, 2016, to review allegations of possible Research Misconduct made against Dr. Mingjun Zhang [REDACTED] ("Respondents") by an anonymous complainant. The Committee of Initial Inquiry (CII) was composed of the following members:

- Dr. Helen Chamberlin (Chair), Professor, Molecular Genetics, College of Arts and Sciences;
- Dr. Martha Belury, Professor, Human Nutrition, College of Education and Human Ecology;
- Dr. Jeffrey Chalmers, Professor, Chemical and Biomolecular Engineering; and
- Dr. Ann Salimbene, Assistant Dean for Administration, College of Engineering.

### Consultant

- Dr. Mike Ibba, Professor and Chair, Department of Microbiology, College of Arts and Sciences

### Ex Officio Members:

- Jennifer K. Yucel, Ph.D., Associate Vice President, Research Compliance/Research Integrity Officer
- Courtney D. Mankowski, Research Integrity and Compliance Manager

Regarding the allegations of possible Research Misconduct involving the intentional misrepresentation (Falsification) of the identity of the newly identified gene for an arabinogalactan protein from English Ivy (*IAGP*) brought against Dr. Mingjun Zhang, Professor, Department of Biomedical Engineering, [REDACTED] the CII originally determined, under the preponderance of evidence standard, by a vote of 3 in favor and 1 against, that the allegations **did not have** sufficient substance to indicate that Research Misconduct may have occurred and therefore should be dismissed<sup>7</sup>.

Following evaluation of the response to the Inquiry report provided by Dr. Zhang, on April 5, 2017<sup>8</sup>, the CII's original determination [REDACTED]. The CII did change its determination regarding Dr. Zhang, finding by a vote of 3 in favor and 1 against, that the allegations **did have** sufficient substance to warrant moving forward to Investigation for Dr. Zhang only<sup>10</sup>.

<sup>5</sup> ATT-5 - 20160823 – Transcript RIO meeting with [REDACTED] Zhang #2

<sup>7</sup> ATT-7 - 20170321 – CII Preliminary Report – Zhang [REDACTED]

<sup>8</sup> ATT-8 - CII A27 – Respondents response to PR (folder)

<sup>9</sup> ATT-9 - 20170428 – CII Final Report – Zhang [REDACTED]

<sup>10</sup> ATT-9 - 20170428 – CII Final Report – Zhang [REDACTED]

On April 28, 2017, the Final report and the CII's letter<sup>11</sup> was provided to Dr. Whitacre for her review, in accordance with the Policy. On May 19, 2017, Dr. Whitacre provided her review to the CII<sup>12</sup>. In her letter, Dr. Whitacre requested that the CII review the evidence and reconsider its findings. Dr. Whitacre granted the CII an additional two (2) month period in order to accomplish their review and reconsideration<sup>13</sup>. The CII re-reviewed a number of pieces of previously reviewed evidence, conducted additional reviews of new evidence obtained from Dr. Zhang's computer, reviewed additional information provided by [REDACTED], and conducted additional interviews.

After the final reconsideration of the evidence, including additional new findings by the CII, the CII ultimately determined, under the preponderance of the evidence standard, by a vote of 3 in favor, and 1 against, that the allegations against [REDACTED] Dr. Zhang [REDACTED] **did have** sufficient substance to indicate possible Research Misconduct, and therefore both warranted Investigation under the Policy. The detailed analysis of the CII's final reconsideration begins on page 17 of the Revised Final Report of the Committee of Initial Inquiry (CII) Concerning Allegations of Research Misconduct, dated July 21, 2017<sup>14</sup>.

Briefly, the change in the CII's determination was primarily driven by Dr. Zhang's response to the CII's Preliminary Report in which Dr. Zhang submitted evidence that he was extensively involved with all aspects of [REDACTED] and overseeing all research conducted in the lab. The information provided by Dr. Zhang in his response to the CII's Preliminary Report was in direct opposition to his earlier statements recorded during the CII's interview on January 11, 2017 and also in many other communications with Dr. Yucel during this process. The conflict in Dr. Zhang's statements raised the issue of credibility of Dr. Zhang's testimony for the CII. Were all of Dr. Zhang's [REDACTED] previous statements and testimony correct, or was the new version put forth in Dr. Zhang's response to the Preliminary Report correct? The new evidence provided in Dr. Zhang's response to the CII Preliminary Report, if taken as fact, would support a conclusion by the CII of willful direction by Dr. Zhang relevant to possible Falsification of data presented in Figure 3 of the PNAS publication. The CII thus determined that there was sufficient evidence to support a possible conclusion of intentional or knowing Falsification of reported research in the PNAS 2016 paper and that the allegations warranted further Investigation.

In summary, based on all of the evidence reviewed and the testimony of Dr. Zhang [REDACTED], the CII affirmed its earlier determination that the data regarding the sequence identity in the PNAS 2016 paper was misrepresented and that the accepted practices of the scientific discipline were ignored or intentionally not followed in the generation, interpretation, and reporting of the identity of the Ivy *IAGP*. With the addition of the new findings, the CII determined that there was sufficient evidence at the preponderance of evidence standard to move [REDACTED] Dr. Zhang [REDACTED] forward to Investigations. Especially compelling for the CII during their reconsideration was 1) the presence of all of the new identified documents on Dr. Zhang's computer, supporting the claim that he was very closely overseeing [REDACTED] in direct conflict with his previous statements;

[REDACTED]

<sup>11</sup> ATT-10 - CII A34 – 20170428 – Letter Chair to Respondents - Reconsideration

<sup>12</sup> ATT-11 - CII A36 – 20170519 – SVPR to Respondents and Chair – Appeal Response

<sup>13</sup> ATT-12 - CII A37 – 20170525- Email SVPR to RIO – approval Deadline extension and approval for external notification - Zhang case

<sup>14</sup> ATT-13 - 20170721 – CII Revised Final Report – Zhang [REDACTED]

## **College of Engineering Investigation Committee**

### **COEIC Voting Members**

- Dr. Joseph P. Heremans (chair), Ohio Eminent Scholar, Department of Mechanical and Aerospace Engineering, College of Engineering,
- Dr. Umit S. Ozkan, COE Distinguished Professor, Department of Chemical and Biomolecular Engineering, College of Engineering; and
- Dr. Gerald Frankel, DNV Chair and Professor, Department of Materials Science and Engineering, College of Engineering.

### **Consultants**

- Dr. Jay Hollick, Associate Professor, Department of Molecular Genetics, College of Arts and Sciences
- Dr. Vicki Wysocki, Ohio Eminent Scholar, Department of Chemistry and Biochemistry, Campus Chemical Instrumentation Center (CCIC) – Director

### **Ex Officio Members:**

- Dr. Jennifer K. Yucel, Ph.D., Associate Vice President, Research Compliance/Research Integrity Officer (RIO)
- Dr. Julia Behnfeldt, Associate Director, Office of Research Compliance/RIO
- Courtney D. Mankowski, Research Integrity and Compliance Manager

## **College of Engineering Investigation Committee Meetings**

September 13, 2017 – Initial meeting with COEIC. Discussed Policy, process and committee charge.<sup>15</sup>

October 9, 2017 – COEIC working meeting. Introduction of Dr. Hollick as expert consultant on molecular biology techniques. Discussion of NSF referral letter<sup>16</sup>, status of Respondent, in depth discussion of molecular biology techniques employed in disputed data.

October 12, 2017 COEIC working meeting. Discussion of the techniques and procedures used in the PNAS paper.

November 2, 2017 COEIC working meeting. Reviews of the original submission documents to Nature Plants<sup>17</sup>, and discussion of the referee reports received from Nature Communications<sup>18</sup>.

November 9, 2017 COEIC working meeting. Preparations of the questions to ask during the interview [REDACTED]

November 21, 2017 – ORC staff meeting with Dr. Vicki Wysocki, Director of CCIC and Faculty Advisor to OSU's Mass Spectrometry facility to act as consultant to the COEIC. Discussed the Policy, process and charge to COEIC as well as the role of a consultant in the process.

November 30 2017 - COEIC working meeting. Introduction of expert consultant Dr. Wysocki; discussion of the origin of the MALDI data; Finalizing the questions to ask [REDACTED]

December 4, 2017 – COEIC interview with [REDACTED]

<sup>15</sup> ATT-14 - 20170913 – COEIC Charge - Zhang

<sup>16</sup> ATT-15 - 20170913 – Investigation Referral Letter for Zhang I-17-0092-O

<sup>17</sup> ATT-16 - CII A14 – Nature Plants Materials and Data provided to COEIC 20171109



December 7, 2017 - COEIC working meeting. Debrief of [REDACTED].

January 11, 2018 - COEIC working meeting. Preparations of the questions to ask during Dr. Zhang's interview.

January 22, 2018 - COEIC working meeting. Finalizing the questions to Dr. Zhang.

January 23, 2018 – COEIC interview with Dr. Mingjun Zhang.

January 30, 2018 - COEIC working meeting. Debrief of Dr. Zhang's interview; discussion of the Zhang- [REDACTED] email exchanges; preparation of the questions for [REDACTED] second interview.

February 6, 2018 - [REDACTED] second interview (Dr. Heremans sole participant from the COEIC).

February 13, 2018: COEIC working meeting. Report of Dr. Heremans to the COEIC about [REDACTED] second interview. Debrief of [REDACTED] second interview.

March 13, 2018 - COEIC working meeting. Analysis of written response from Dr. Li Tan<sup>19</sup>; identification of the need for a second interview with Dr. Zhang.

March 23, 2018 – COEIC second interview with Dr. Mingjun Zhang.

March 30, 2018 - COEIC working meeting. Debrief second interview with Dr. Zhang.

April 20, 2018 – COEIC meeting to finalize determination and vote on allegations.

July 23, 2018 - COEIC meeting to discuss response from Dr. Zhang and finalize determination and report.

## Research Records and Evidence

- Folder of materials provided to Committee November 9, 2017 regarding Nature Plants submission<sup>20</sup>
- Folder of materials provided to Committee November 14, 2017 with data regarding weekly reports from [REDACTED] Zhang<sup>21</sup>
- Folder of materials provided to Committee November 28, 2017 with data regarding N-terminal sequencing<sup>22</sup>
- Folder of materials provide to Committee November 30, 2017 with data regarding PNAS submission, Nature Communication submissions, sequence data<sup>23</sup>
- [REDACTED]
- Audio recordings and transcripts for RIO meetings with Dr. Zhang [REDACTED] conducted on August 23, 25, 2016, provided to Committee December 28, 2017<sup>25</sup>
- Folder of materials provided to Committee January 18, 2018 with data regarding weekly reports and project updates<sup>26</sup>
- Folder of materials provided by Dr. Zhang January 22, 2018 in advance of his interview with COEIC<sup>27</sup>
- Folder of materials provided to Committee January 30, 2018 regarding manuscript submissions for various journals<sup>28</sup>

<sup>19</sup> ATT-25 - 20180301 - Response from Li Tan

<sup>20</sup> ATT-16 - CII A14 – Nature Plants Materials and Data provided to COEIC 20171109

<sup>21</sup> ATT-17 - 20171114 – Data provided to COEIC

<sup>22</sup> ATT-18 - 20171128 – Data provided to COEIC

<sup>23</sup> ATT-19 - 20171130 – Data provided to COEIC

<sup>25</sup> ATT-21 - Audio recordings August 2016

<sup>26</sup> ATT-22 - 20180118 – Data provided to COEIC

<sup>27</sup> ATT-23 - 20180122 – Data provided to COEIC

<sup>28</sup> ATT-24 - 20180130 – Data provided to COEIC



- Response Letter from Li Tan dated March 1, 2018<sup>29</sup>
- July 3, 2015 Email communication [redacted] Zhang about removing sequence<sup>30</sup>

[redacted]

- COEIC Interview with Dr. Zhang, January 23, 2018 and named exhibits<sup>32</sup>
  - Exhibit 1 – Slide from /Jarvis/utk/drug delivery/condensed slides 3013-4/condensedIvyNPUpdate 2013-5-15
  - Exhibit 2 – Slide from /Jarvis/utk/drug delivery/condensed slides 3013-4/condensedIvyNPUpdate 2013-5-16
  - Exhibit 3 – [redacted] 2016 PNAS publication
  - Exhibit 4 –N-sequencing result 13-5-30
  - Exhibit 5 –result 2013-3-25
  - Exhibit 6 –20171213 - Email [redacted] RIO - Fwd\_ Fw\_ interest in neal stewart's Hedera helix sequences
  - Exhibit 7 - Reviewers comments on 2015 Nature Communications manuscript
  - Exhibit 8- 20180118 Manuscript Submission Dates
  - Exhibit 9- Supplemental Information to [redacted] 2016 PNAS
  - Exhibit 10- Attachment 41 from the CII – First round of sequence
  - Exhibit 11- Neal\_Final\_Email\_End\_Collaboration\_June\_2013
  - Exhibit 12- Transcriptome blast based on source data of ground root 13-8-6 (internal blast by Eric).docx
  - Exhibit 13- 5' RACE.docx
  - Exhibit 14- From NCBI Genbank – Record KF752597
  - Exhibit 15-From NCBI Genbank – Record KM820289
  - Exhibit 16- Response to the Allegations against the PNAS paper.pdf
  - Exhibit 17- Nature Communication Submission- "communications with PNAS and Nature Communications 90-140"

[redacted]

- Interview with Dr. Zhang, March 23, 2018 and named exhibits<sup>34</sup>

<sup>29</sup> ATT-25 - 20180301 – Response from Li Tan

<sup>30</sup> ATT-26 - Emails July 3, 2015- Fwd\_Decision on manuscript NCOMMS-14-00999A-Z

<sup>32</sup> ATT-28 - 20180123 – COEIC Interview Zhang #1 and exhibits

<sup>34</sup> ATT-30 - 20180323 – COEIC Interview Zhang #2 and exhibits



- Exhibit 1 - COEIC Interview with Dr. Zhang, Jan 23, 2018
- Exhibit 1a - 20180123 – COEIC Interview Zhang pg 44-45
- Exhibit 1b - 20180123 – COEIC Interview Zhang pg 43
- Exhibit 1c - 20180123 – COEIC Interview Zhang pg 54
- Exhibit 1d - 20180123 – COEIC Interview Zhang pg 21
- Exhibit 1e - 20180123 – COEIC Interview Zhang pg 23
- Exhibit 1f - 20180123 – COEIC Interview Zhang pg 43 44 47 66
- [REDACTED]
- [REDACTED]
- Exhibit 3 - Emails July 3, 2015- Fwd\_Decision on manuscript NCOMMS-14-00999A-Z
- Exhibit 4 - Communications from Li Tan
- [REDACTED]
- Exhibit 6 - 20170327 – Letter of Support for Zhang [REDACTED] Yucel
- Exhibit 7 - Submission and Experimental Timeline
- MZ Exhibit 1 – “Summary” document (hard copy brought to interview)
- MZ Exhibit 2 – “Explanation to E-mails between [REDACTED] and Me” (hard copy brought to interview)
- MZ Exhibit 3 – Series of emails from Li Tan to Zhang (hard copy brought to interview)
- Email from Dr. Zhang to Yucel, March 28, 2018 with 6 attachments and many additional files loaded to Box. All placed in a folder named 20180328 – Documents from Zhang<sup>35</sup>.

### **Submission and Experimental Timeline**

From data provided by [REDACTED] Dr. Zhang as well as documents found on the various computers, the Committee generated the following timeline to highlight key experimental results as they relate in time to the various submissions of some form of the manuscript.

#### **2013**

- Nature Nanotechnology submission (1/7/13, date unclear, figures have 2014 file date)
- Nature Communications submission (1/20/13, date on letter, but is likely 2014 by file date)
- Protein/peptide analysis report obtained from Iowa State Univ (3/22/13)
- Protein /peptide analysis report obtained from Iowa State Univ (5/29/13)
- 3' RACE DNA sequence ('first round of sequence') obtained from McLab (9/4/2013)
- Partial mRNA sequence of the putative Ivy AGP was submitted to Genbank (10/23/13)
- Science submission (10/28/13)
- Nature submission (11/25/13)
- Mass spec data acquired at UTK facility (12/11/13)
- Nature Chem Biology submission (12/12/13)
- 5' RACE DNA sequence (5RACE Seq) obtained from McLab (12/15/2013)

#### **2014**

- Genomic DNA sequence obtained from McLab (1/13/14)
- Nature Materials submission (6/11/14)
- Dr. Li Tan, Univ of Georgia joined project (7/3/2014)
- Nature Plants submission (8/7/14)
- Nature Plants (revision) (Sept. 2014)

<sup>35</sup> ATT-31 - 20180328 – Documents from Zhang





- Composite sequence (5' and 3' RACE sequence plus genomic intron) submitted to GenBank (9/29/14)
- Science submission (11/8/14)

**2015**

- Nature Biotechnology submission (1/11/15)
- Nano submission (1/22/15)
- First PNAS submission (1/25/15)
- Nature submission (4/6/15)
- Current Biology submission (4/8/15)
- PLoS Biology submission (4/10/15)
- Nature Communications submission (4/18/15)
- Mass spec data acquired at OSU CCIC facility (9/9/15)

**2016**

- Second PNAS submission (1/12/16)
- Accepted at PNAS (4/29/16)
- Published in PNAS (5/23/16)

**Interview Summaries**

During the course of the Investigation, the COEIC conducted four (4) interviews. [REDACTED]

[REDACTED] interviewed Dr. Zhang on January 23, 2018<sup>38</sup> and again on March 23, 2018<sup>39</sup>. All of the interviews were recorded and transcribed and are provided with this report.

In general, while the COEIC found [REDACTED] Dr. Zhang generally cooperative with the investigation and willing to provide requested information and materials, the COEIC notes [REDACTED] This made it difficult for the COEIC to know what the truth is and to determine the credibility of either person. During earlier phases of the process, [REDACTED] Dr. Zhang stated very firmly that there was nothing wrong with the sequence that was identified and published. The COEIC notes [REDACTED]

The COEIC committee asked Dr. Zhang [REDACTED] had ever expressed concerns about the sequence or indicated to him that they should remove the DNA sequence. Dr. Zhang was adamant that [REDACTED] never told him that there were problems with the sequence [REDACTED] had never indicated that the sequence should be removed<sup>41</sup>. When asked by the COEIC about the July 3, 2015 email exchange between Dr. Zhang [REDACTED] Dr. Zhang indicated that he did not remember the email, that he probably was not recognizing that [REDACTED]

<sup>38</sup> ATT-28 20180123 – COEIC Interview – Zhang #1

<sup>39</sup> ATT-30 - 20180323 – COEIC Second Interview – Zhang #2

<sup>41</sup> ATT-30 - Zhang 3/23/18 Transcript – pg 14 ln 20 through pg 15, ln 41 pg 16 ln 21-24; pg 17 ln 1-4

<sup>42</sup> ATT-26 - Emails July 3, 2015 – Fwd\_Decision on manuscript NCOMMS-14-00999A-Z



██████████ was raising concerns about the sequence<sup>43</sup>. The COEIC notes that miscommunication between Dr. Zhang ██████████ could have had a large influence on the course of their actions.

Additional specific responses from Dr. Zhang ██████████ are included below in the Investigation Committee Analysis section of the report.

### **Notification to Respondent –Information Release**

On December 12, 2017<sup>44</sup>, Dr. Zhang was notified of the Institution's intent to contact the following individuals with questions related to their role on the project or to obtain corroboration of statements made by Dr. Zhang ██████████ related to the PNAS 2016 paper. The external contacts were by letter to:

- Dr. Eric Carpenter, Database Manager, University of Alberta regarding availability and access to sequence data for the 1KP project<sup>45</sup>. Dr. Yucel attempted to contact Dr. Carpenter by email on January 24, 2018 and made a second request on February 7, 2018. No response was received.
- Dr. Yongzhong Wang (coauthor), PolyCreative Health Biotech Co. Ltd, China requesting that Dr. Wang participate in an interview with the committee(s)<sup>46</sup>. Dr. Yucel attempted to contact Dr. Wang by email on December 12, 2017, made a second request on December 29, 2017 and a final request on March 26, 2018 (email sent to his university email address in addition to the company address previously tried). No response was received.
- McLab, South San Francisco, CA requesting information regarding the sequencing services they provided related to the PNAS paper<sup>47</sup>. Dr. Yucel attempted to contact the vendor by email on December 12, 2017, and made a second request on December 29, 2017. No response was received.

As required under the Policy, the then Senior Vice President for Research, Dr. Caroline Whitacre, reviewed and authorized these contacts by the Office of Research Compliance on December 10, 2017<sup>48</sup>.

On February 26, 2018, Dr. Zhang was notified of the Institution's intent to contact Dr. Li Tan, at the University of Georgia, with questions related to his role on the project<sup>49</sup>. Dr. Li Tan was a collaborator on the project and listed as a co-author on the PNAS 2016 paper. As required under the Policy, the Interim Senior Vice President for Research, Dr. Randolph Moses reviewed and authorized this contact by the Office of Research Compliance on February 22, 2018<sup>50</sup>. The Office of Research Compliance sent a letter to Dr. Tan with a number of questions that both investigation committees had related to this matter<sup>51</sup>. Dr. Tan provided his response to the Committee's questions to Dr. Yucel on March 1, 2018<sup>52</sup>.

<sup>43</sup> ATT-30 - Zhang 3/23/18 Transcript – pg 23 ln 12 through pg 24 ln 9

<sup>44</sup> ATT-32 - 20171212 – Email RIO to Respondents – Notification of external contacts

<sup>45</sup> ATT-33 - 20170124 – Letter RIO to Carpenter – 1KP

<sup>46</sup> ATT-34 - 20171212 – Letter RIO to Wang

<sup>47</sup> ATT-35 - 20171212 – Letter RIOs to Mclab

<sup>48</sup> ATT-36 - 20171210 – SVPR approval for external contacts

<sup>49</sup> ATT-37 - 20180226 – RIO to Zhang ██████████ – Notify external contact Tan

<sup>50</sup> ATT-38 - 20180226 – ISVPR Approval for external notification

<sup>51</sup> ATT-39 - 20180226 – Letter RIO to Tan – Investigation committee questions and attachment

<sup>52</sup> ATT-25 - 20180301 – Response from Li Tan



## **Investigation Committee Analysis**

**Manuscript #1 - Allegation #1** - The identity and species of origin of the Ivy arabinogalactan protein (IAGP) was falsely reported by Dr. Zhang by the intentional, knowing, and/or reckless misrepresentation of data, text, and figures related to the isolation and characterization of the clone identified, published in Fig. 3 and Fig. S4A of PNAS 113(23):E3193-3202 and deposited in GenBank database as (KM820289).

## **Finding of Fact**

1. The following are instances the Committee determined to be knowing and reckless text misrepresentations of identity and origin of the IAGP in PNAS paper:
  - a. Physicochemical analyses support this identification but the specific arabinogalactan protein identified in this article is not secreted by English Ivy, given comment (b) and the analysis provided in (c) hereunder.
  - b. The Nature Communication reviewer of a previous version of this manuscript made the authors aware of this concern with the species identification of the published IAGP sequence.
  - c. On page E3195, the authors report:
    - i. “For such a purpose, the AGP-rich fraction gathered from RP-HPLC (i.e., fraction 4) was deglycosylated.” This is not a true statement as:
      1. Per the information noted on the Iowa State submission form, the samples submitted to the Iowa State University Protein Facility were not deglycosylated according to the written information on the submission form<sup>53</sup>.
      2. Further, the peptide sequence obtained from the two samples submitted to the Iowa State University Protein Facility are inconsistent with the sample consisting of a single protein species.
    - ii. “and a short segment of an amino acid sequence at the N terminus of the deglycosylated protein, Ala-Hyp-Hyp-Hyp-Thr-Asp-Ala, was determined via Edman degradation.” This is not a true statement as:
      1. No such peptide sequence could be determined to be in Ivy nanoparticles.
      2. The peptide sequence APPPTDA could not have been determined from the Edman degradation results from Iowa State University.
    - iii. The amino acid sequence APPPTDA matches that encoded by the cDNA sequence they obtained from MCLab and the cDNA sequence provided by MCLab was based on nucleic acid sequence obtained from OneKP project information.
    - iv. “and the nucleotide sequence for the full-length cDNA encoding the core protein was then determined by 5'- and 3'- RACE cloning (Fig. 3B).” This is not a true statement as:
      1. It is a clear deviation from accepted practice to piece together two nucleic acid sequences at a common overlap (the OneKP primer sequence) and conclude where the hypothetical encoded protein begins without additional supporting information. In fact, if one assumes that the two pieces (the 5' and 3' RACE sequences) are contiguous in nature, then someone with even rudimentary knowledge of nucleic acid biology would recognize that a much larger protein could be encoded. The disregard for all the protein coding potential of the 5'

---

<sup>53</sup> See ATT-28, exhibits 4 and 5



RACE sequence in addition to not reporting the entire 5' RACE sequence in some format indicates a willfulness to deceive.

- v. “Moreover, the identified full protein sequence, designated as IVY ARABINOGALACTAN PROTEIN (IAGP), demonstrated a moderate similarity to four other HRGPs derived from Arabidopsis in a sequence alignment (Fig. 3C).” This is a misleading statement possibly influenced by fabricated data as:
  1. Qualitative/subjective statements like this constitute a significant departure from accepted practice. Because the protein alignments were made manually, the author is free to place gaps that maximize similarities.
  2. A casual reader would use the shading of identical (dark) or similar (light) amino acid residues as a visual assessment of identity/similarity.
  3. The number of dark shaded residues in particular makes a compelling visual argument for or against such protein relatedness. Since reviewer 4 of the Nat Communications manuscript specifically pointed out that the 10th amino acid residue of the "IAGP" was incorrectly shaded dark, we would consider the failure to correct this a deliberate falsification to bolster what is in fact a highly subjective claim.
  
- vi. “Notably, in a BLASTp search in the genome of the *A. thaliana*, the putative IAGP identified from *H. helix* also exhibited a moderate extent of similarity to several cytochrome c oxidase subunits” This is a false statement as:
  1. a BLASTp search confined to the *A. thaliana* genome would not return such a finding.
  2. A reiterative BLASTp search (Psi-Blast) does identify these cytochrome c oxidase subunits, but at a level of confidence below that considered acceptable for definitive identification.
  
- vii. “... implying potential homology between these two types of proteins.” This is a misleading statement as:
  1. By confining the BLAST searches to the *A. thaliana* genome, the authors did not follow standard molecular biology protocol and thus misled the reader by implying that *A. thaliana* proteins are the most closely related to the presumed IAGP.
  2. Reviewer 4 of the Nature Communications submission (4/18/15)<sup>54</sup> identified that the presumed IAGP protein was most closely related to fungal proteins.
  
- viii. “Accordingly, a phylogenetic analysis was carried out to determine the relationship of the IAGP with these analogous proteins obtained from BLASTp.” This is an intentionally false statement as:
  1. the other Arabidopsis protein sequences aligned in Fig 3C could not have been identified in a BLASTp search as specified.
  
- ix. “Among them, the IAGP and the AGP9 were the closest relatives, as shown in Fig. S4A.” This is a misleading statement as:

---

<sup>54</sup> ATT-26 - Emails July 3, 2015 – Fwd\_Decision on manuscript NCOMMS-14-00999A-Z



1. While this is a technically accurate statement according to the published method of analysis, confining the analysis to just *A. thaliana* proteins fails to identify the most closely-related proteins, which are from fungi.
- d. On page 2/11 in the Supplementary Information, Protein Backbone Analysis section, the authors report:
  - i. “N-terminal sequencing of the deglycosylated AGPs...”: this statement contradicts the information on the Iowa State submission form [REDACTED] indicated that that proteins were ‘heavily glycosylated’. This information must be considered in the light of the contradictory fact that [REDACTED] indicated that the protein was deglycosylated.
  - ii. “According to the obtained N-terminal sequence”. This is a false statement as, the N-terminal sequence could not have been used as a reference for the indicated degenerate primer.
  - iii. “5'-GCICCCICCCIACIGAT/CGCI-3' (I = inositol), was used for a PCR amplification.” This is a false statement as: This sequence is based on the MCLab F3 sequence, one of five 3' cDNA clones obtained using PCR with defined oligonucleotide primer "F2", which is identical to the OneKP sequence provided to MCLab [REDACTED]
  - iv. “ followed by sequencing (MCLAB).” This is a misleading statement as: MCLab performed both the 5' and 3' RACE cloning and the sequencing.
  - v. “ Full-length gene was obtained by amplifying the genomic DNA template using a pair of primers comprising IIP1, 5'-ACCGCTGGGTTTCGGTTACC-3, and IIP2, 5'-TCACGCCCTGGGTAACACG- 3'” This is a misleading statement as:
    1. The full-length "gene" includes all sequences that are converted into a primary RNA molecule including all intron sequences. The authors cannot claim to know what the full gene sequence is unless all the regions are amplified and sequenced from genomic DNA. Not only have they not evaluated whether or not there is intronic sequences upstream of the IIP1 primer sequence, but they have not validated where the 5' transcriptional start site is located.
  - vi. “Similar HRGPs sequences in *A. thaliana* were searched by BLAST,” This is a misleading/false statement as:
    1. Arabidopsis HRGP proteins cannot be identified using BLAST, BLASTp, or reiterative BLAST searches.
2. Reckless misrepresentations of sequence data in Figures
  - a. Figure 3B, the authors reported false statements/data representations in the figure as follows:
    - i. The identification of the underlined amino acid sequence is based on the 5' RACE clone DNA sequence. Given that the 5' RACE clone has the potential to encode far more protein sequence (which is identical to that of fungal RAS-GTPases) than what is indicated here, this depiction of "full-length" protein sequence is without confirmation or supporting information.
    - ii. The figure as published represents a presumed genomic DNA sequence as evidenced by the presence of intronic sequence (dashed lines position 144-227) and not a cDNA sequence, as claimed in the text of the paper, which would only have exonic sequences.
    - iii. The text on page E3195 states “... and the nucleotide sequence for the full-length cDNA encoding the core protein was then determined by 5'- and 3'- RACE cloning (Fig. 3B).”



The figure represents a protein translation of isoleucine (“I”) for an AUG code at amino acid position 48, which should be methionine (“M”).

- b. In the Figure legend for Figure 3B, the authors reported false or misleading statements/data representations as follows:
    - i. The authors reported “(B) Amino acid sequence and corresponding nucleotide sequence of the protein backbone in the AGP-rich ivy nanoparticles”, when this is not a proven statement.
    - ii. The authors reported in the Figure legend “(B) Amino acid sequence and corresponding nucleotide sequence of the protein backbone in the AGP-rich ivy nanoparticles. The amino acids obtained from N-terminal sequencing are single underlined.” (Indicated in figure by single underlined amino acid sequence ‘AOOODA’.) This is not a true statement [REDACTED] did not obtain that full sequence from N-terminal sequencing and the ‘O’s were determined by MALDI-TOF and not N-terminal sequencing.
  - c. In Figure 3C, the authors reported false statements/data representations in the figure as follows:
    - i. The figure falsely represents the 10th amino acid residue (“A”) as being ‘Identical’ at that position by shading it dark, when it would need to be a (“P”) to be identical to others in the alignment. This was pointed by a previous reviewer and was ignored by the authors.
    - ii. The authors reported in the figure legend for 3C “... These four glycoproteins were chosen for multiple sequence alignment because they are the most analogous HRGPs to the IAGP, according to a BLASTp search in *A. thaliana*.” This is a false statement and was not how or why these other sequences were selected for alignment.
3. Knowing and Reckless misrepresentation of sequence data in a national database:
- a. Dr. Zhang allowed the submission of a partial cDNA sequence to the GenBank database (accession no. KF752597) identified as belonging to *Hedera helix* without proper validation.
  - b. Dr. Zhang allowed the submission of sequence purportedly representing a *Hedera helix* gene encoding IAGP to the GenBank database (accession no. KM820289), knowing that it was not complete or correct and left it in GenBank even after serious concerns were raised in subsequent journal reviews.

### **Knowledge and Intent**

The committee finds a large number of misrepresentations that collectively result in the falsification of the identity of the arabinogalactan protein. The committee also finds that Dr. Zhang was well aware of the importance of that identification and its impact on the probability that the work be accepted in a high-impact publication (see the section below on significance). A further motivation for Dr. Zhang’s actions in this matter may have stemmed from growing frustration with trying to get the project published in light of the frequently negative reaction of journal editors and referees following the very large number of unsuccessful previous submissions. From information gathered during this case, there appears to have been at least sixteen (16) previous submissions of the manuscript before the final submission and acceptance at PNAS (see Timeline section above).

Dr. Zhang testified that he was unfamiliar with the techniques of molecular biology used to determine the identity of the arabinogalactan protein. The committee discussed at length the possibility that Dr. Zhang may have been misled unknowingly [REDACTED]. However, the risks and ambiguities associated with the preparation of the samples and the interpretation of the data were made clear to Dr. Zhang repeatedly as documented in this report, and most strongly so in the report



of referee 4 of the submission to Nature Communications (4/18/15). Dr. Zhang, as corresponding author of the manuscript, and as PI of the grant that funded this work, had to be aware of this risk. In his testimony, Dr. Zhang stated that he instructed the other authors to consider the comments of referee 4 and that they convinced him that their interpretation was correct, but he was unable to provide evidence, such as email records, to support this claim. [REDACTED] The commonly accepted standard in the scientific community is that the PI and corresponding author is responsible for the accuracy of the reported results.

Given the strongly worded comments of referee 4 of the Nature Communications (4/18/15) submission, the convincing evidence of the validity of those comments, and the absence of response to those comments either in the text of the paper subsequently submitted to PNAS or in internal deliberations, it appears that Dr. Zhang consciously chose to disregard the comments and deliberately resubmitted the manuscript to PNAS without any further validation or justification. Excerpts of the comment are reproduced here:

*“Reviewer #4 (Remarks to the Author)<sup>55</sup>*

*This manuscript describes an intriguing hypothesis that a calcium dependent interaction between arabinogalactan and pectin provides a mechanism for adhesion of ivy. If true, this would be of considerable interest, given the poor understanding of arabinogalactan mechanism and function, and also given interest in plant adhesives and cell adhesion. Unfortunately, as I studied the experiments in this paper to understand fully the evidence, it became clear to me that many of them are very poorly conducted, and very much over interpreted. The experiments are sometimes not as they are described, they are redundant, or they are uninformative. I have therefore not reviewed all aspects of the manuscript.*

*A few examples will suffice.*

*(...)*

*P7, line 12: "a full-length cDNA was amplified and cloned (Fig. 3a)."*

*However, the sequence in Fig3a is genomic sequence, as it contains an intron, as described in the legend. Why show the genomic sequence?*

*(...)*

*9. There being almost nothing correct in this figure, I searched the amino acid sequence against NCBI. It is identical to a fungal cytochrome C oxidase subunit. The authors have not identified an ivy AGP. They have cloned a fungal protein.*

*On the basis of this figure alone, the manuscript shows that the authors have not taken due care and attention to their experimentation, or to manuscript preparation.”*

## **Respondent’s Response**

Dr. Zhang has repeatedly made the following points in his defense. He has argued, [REDACTED], that he had a lack of expertise in the subject of molecular biology.<sup>56</sup> He singled out his lack of expertise on the subject of N-Terminal sequencing and deglycosylation,<sup>57</sup> and of cDNA sequencing.<sup>58</sup> The

<sup>55</sup> ATT-26 - Emails July 3, 2015 – Fwd\_Decision on manuscript NCOMMS-14-00999A-Z

<sup>56</sup> ATT-30 - Zhang 3/23 transcript pg 17, ln 11-17 and pg 70 ln 1-8 (use of BLAST)

[REDACTED]  
TT-28 - Zhang 1/23 transcript pg 67 ln 16 through pg 69 ln 20; pg 71 lns 3-21; pg 72 lns 13-20; pg 73 lns 7-12  
ATT-30 - Zhang 3/23 transcript pg 38 ln 6 – pg 39 ln 19 (Li Tan, not glycosylated before N term seq); pg 39 ln 19, pg 40 ln 13-pg 41 ln 24 (Li Tan seq concerns)

[REDACTED]



committee notes that Dr. Zhang also expressed ignorance about the details of the 1KP project and the McLab sequence results.<sup>59</sup> The subject matter of the referee #4 report mentioned above was the object of extensive questioning of [REDACTED] Dr. Zhang. The answers followed two main themes, frustration with the review procedures and miscommunications [REDACTED] Dr. Zhang.<sup>60</sup> The outcome of these discussions were often contradictory<sup>61</sup> [REDACTED] The responsibility of the corresponding author, Dr. Zhang, was discussed.<sup>62</sup> It was unclear that Dr. Zhang accepted the responsibility for submitting the data to Genbank.<sup>63</sup> The totality of this information was taken into account in the committee’s final vote.

**Significance**

Dr. Zhang’s assessment of the importance of the work is evident by the number and profile of the journals to which Dr. Zhang had previously submitted the paper (see Timeline section above). It is very likely that without an identified sequence for IAGP, the manuscript would not have been accepted by PNAS, thus necessitating the continued inclusion of the DNA sequence in the manuscript despite concerns raised by reviewers. While at times during the Investigation, Dr. Zhang indicated that he believed that the sequence of the protein was not important or critical to the manuscript,<sup>64</sup> this is contradicted by the fact that Dr. Zhang was quoted in a May 2016 press release<sup>65</sup> for the PNAS paper stating: “By understanding the proteins that give rise to the high strength of Ivy’s adhesive, we can inspire approaches to engineer and synthesize new bio-inspired adhesives for medical and industry products,” indicating the importance of identifying the protein sequence.

[REDACTED] ATT-28 - Zhang 1/23 transcript pg 44 ln 12-18 (anyone suggest removing seq); pg 45 ln 13-18 [REDACTED] pg 110 ln 4 through pg 111 ln 12; pg 46 ln 4-12 (significance of Fig 3 to paper); ATT-30 - Zhang 3/23 transcript pg 23 ln 12 – pg 24 ln 9 (July 3 email discussion)

[REDACTED] - Zhang 1/23 transcript pg 78 ln 10 through pg 81 ln 11; pg 82 ln 15-18; pg 83 ln 9-23

[REDACTED]

[REDACTED]

<sup>62</sup> ATT-28 - Zhang 1/23 transcript pg 50 ln 10-19; pg 112 ln 5 through pg 113 ln 3

[REDACTED]

<sup>64</sup> ATT-30 - 20180323 –COEIC Interview – Zhang #2, pg 27 ln 3-21

<sup>65</sup> ATT-40 - 20160523 – Zhang Press Release – PNAS paper



Furthermore, any thoughtful response to the comments of referee 4 would include performing a BLAST analysis of the DNA sequence. When Dr. Zhang was asked if he knows how to BLAST a sequence and if he BLASTed the sequence published in the paper, he responded that he learned about BLASTing as a student but did not BLAST this sequence, and did not have the time to do so.<sup>66</sup> However, nowadays BLASTing is very easy to do by pasting a sequence into a web app. It is evident to the committee that a deep knowledge of molecular biology is not needed to understand that the BLAST results of the disputed sequence show a high match to fungal species.

In its review of the allegations, the COEIC used the National Science Foundation’s (NSF) Office of Inspector General definition of “Reckless”<sup>67</sup> when making the finding of Research Misconduct. As NSF funded the research under review, their standard is the applicable one. Specifically, NSF defines “Reckless” as *“The subject did not exercise the care a reasonable person similarly situated would have exercised under the circumstances, and did so with a conscious awareness of, or indifference to, the risk of adverse consequences of his actions and the potential resulting harm. Reckless is essentially synonymous with grossly negligent.”*

Dr. Zhang’s actions and response to the questions were found by a majority of the committee to deliberately disregard the critical comments received from referee 4. Dr. Zhang subsequently submitted the manuscript to PNAS without justification for the findings critiqued in the previous submission to Nature Communications. The combination of these actions were therefore found by a majority of the committee to be reckless, based on the facts that Dr. Zhang gave evidence that he understood the importance of the identification of the proteins, was made aware in no uncertain terms that it was wrong, and then knowingly disregarded this information when he submitted the manuscript, without addressing the points raised.

A majority of the Committee determined that the reckless misrepresentation of the species of origin and the sequence identified as encoding *IAGP* and reported by Dr. Zhang does impact the reported results. While the sequence data is only one part of the PNAS paper, the identification and cloning of the putative gene for *IAGP* is central to the reported results. Given the extensive problems with the sequence and the fact that it is most likely not a protein isolated from ivy, it is very important that the sequence be retracted from the scientific record unless or until its authenticity as a protein from ivy can be confirmed.

One member of the Committee felt that, while there were clearly errors in the PNAS paper, and that Dr. Zhang’s actions were extremely careless, they did not rise to the level of “Reckless” required for a finding of Research Misconduct. This Committee member noted the following points:

- 
- 

<sup>66</sup> ATT-30 - 20180323 –COEIC Interview – Zhang #2 pg 70 ln 5

<sup>67</sup> ATT-54 - NSF OIG Assessing Intent in Research Misconduct Investigations - [nsf.gov/oig/outreach/RM-intent.pdf](https://www.nsf.gov/oig/outreach/RM-intent.pdf), pages 2-3

<sup>68</sup> ATT-28 – 20180123 - COEIC Interview Zhang #1 – see , Exhibit 11- “Neal\_Final\_Email\_End\_Collaboration\_June\_2013”



[REDACTED]

[REDACTED]

- In regards to the collaboration started with Dr. Li Tan at the University of Georgia in July 2014, the Committee noted that [REDACTED] with Dr. Tan (in the time frame of 7/3-7/24/2014) however, Dr. Zhang was not copied on any of them.<sup>72</sup>
- [REDACTED]
- [REDACTED]
- When the second rejection from the journal Nature Plants was received, there were direct email exchanges [REDACTED] Dr. Li Tan, who is the expert in the area in which Dr. Zhang is missing the background. There was no evidence presented that Dr. Tan ever raised any concerns about the sequence in question. In fact, even when Office of Research Compliance reached out to Dr. Tan in February 2018, Dr. Tan still did not appear to doubt the integrity of the data presented in the PNAS paper<sup>76</sup>. All of this may have given Dr. Zhang confidence that there was no problem with the sequence data presented and the objections raised by the reviewers were merely differences in opinion or a reluctance to admit a new comer to a field.
- The PNAS paper was reviewed by two sets of reviewers for publication in PNAS, a high impact factor journal. The paper was then accepted for publication. Given that PNAS accepted the paper, Dr. Zhang may have been led to believe that all problems with data presented in the paper had been adequately addressed. The Committee feels that this situation demonstrates a failure in the peer review process at PNAS. The Editor should have required the authors to provide clear and convincing experimental evidence of the identity of the clone and not allowed the authors to restrict their representations to Arabidopsis comparisons only.

[REDACTED]

<sup>71</sup> ATT-25 - 20180301 – Response from Li Tan

[REDACTED]

<sup>76</sup> ATT-25 - 20180301 – Response from Li Tan



- There was little written evidence provided by either Dr. Zhang [REDACTED] [REDACTED] [REDACTED] raising questions about the validity of the data, except an email exchange provided by [REDACTED] [REDACTED] [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

- [REDACTED]

- [REDACTED]

Therefore, it is the opinion of the dissenting committee member that Dr. Zhang was deceived and misled [REDACTED]

[REDACTED] This finding does not excuse the poor supervision and oversight provided by Dr. Zhang, but it does constitute the basis for the “No” vote of this committee member regarding Allegation #1.

**Committee Conclusion - Allegation #1**

By a preponderance of the evidence, the Committee finds by a vote of 2 in favor to 1 against, that the Respondent, Dr. Zhang recklessly falsified and misrepresented the origin and sequence data associated with *IAGP* that he reported in the PNAS 2016 paper, and that this act does constitute Research Misconduct (Falsification), as described in the Policy IIIA and 45 C.F.R. § 689.2(c)(2).

By clear and convincing evidence, the Committee finds by a vote of 2 in favor to 1 against, that the Respondent, Dr. Zhang recklessly falsified and misrepresented the origin and sequence data associated with *IAGP* that he reported in the PNAS 2016 paper, and that this act does constitute Research Misconduct (Falsification), as described in the Policy IIIA and 45 C.F.R. § 689.2(c)(2).

<sup>77</sup> ATT-26 - Emails July 3, 2015 – Fwd\_Decision on manuscript NCOMMS-14-00999A-Z

[REDACTED]

**Manuscript #1 - Allegation #2** - MALDI-TOF mass spectrometry data was falsely reported by the intentional, knowing or reckless misrepresentation of the published methods used including the location where the MALDI-TOF mass spectrometry experiments were conducted in the text and Fig. S4B of PNAS 113(23):E3193-3202.

### **Finding of Fact**

1. In the PNAS 2016 Supplemental Information, on page 2, Dr. Zhang reports “MALDI-TOF MS analyses were carried out at the Mass Spectrometry and Proteomics Facility of The Ohio State University...” The Committee determined that this was a false statement [REDACTED] the published mass spec data was done at the University of Tennessee, Knoxville.
2. In the PNAS 2016 Supplemental Information, on page 2, Dr. Zhang reports “... using a ultrafleXtreme mass spectrometer (Bruker) equipped with 355-nm nitrogen lasers (20Hz)...” The Committee believes that this is most likely a false statement [REDACTED] the published mass spec data was done at the University of Tennessee, Knoxville, and it is unlikely that the text reflects the equipment used at UTK.
3. In the PNAS 2016 Supplemental Information, on page 2, Dr. Zhang reports “... according to the procedure described previously (91)...” The Committee believes that this is a false statement [REDACTED] the published mass spec data was done at the University of Tennessee, Knoxville, and it does not reflect the methods used to prep and analysis the sample done at UTK.
4. In the PNAS 2016 Supplemental Information, on page 2, Dr. Zhang reports “Reflectron ion mode was used for positive ion detection, with ion source voltage of 25 kV and a 90-ns delay.” The Committee believes that this is a false statement [REDACTED] the published mass spec data was done at the University of Tennessee, Knoxville, and it does not reflect the methods used to prep and analysis the sample done at UTK.

### **Knowledge and Intent**

The Committee asked [REDACTED]

[REDACTED] it is unclear if Dr. Zhang knew or understood that there was a discrepancy between the MALDI-TOF data used and the reported methods<sup>81</sup>.

<sup>79</sup> ATT-42 - CII A42 – 31214 IVY-NP

<sup>80</sup> ATT-43 - CII A43 – UTK IVP NP #1 mass spec trace



**Respondent's Response**

The origin and significance of the mass spec data were the object of several questions during the interviews of of Dr. Zhang

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

---

[REDACTED]



[REDACTED]

[REDACTED]

[REDACTED]

### **Significance**

The Committee can find no obvious reason for Dr. Zhang to have intentionally misrepresented where or how the MALDI-TOF data were obtained. The Committee determined that it is not clear that Dr. Zhang was informed of the fact that the data being presented in the PNAS 2016 paper was not obtained at OSU CCIC. The Committee believes that the misrepresentation in the paper is primarily due to carelessness [REDACTED] in the preparation of the manuscript. The Committee has determined that the misrepresentation of the MALDI-TOF data outlined in Allegation 2 by does not impact the reported results.

### **Committee Conclusion - Allegation #2**

By a preponderance of the evidence, the Committee finds by a vote of 0 in favor and 3 against, that the Respondent, Dr. Zhang intentionally, knowingly or recklessly reported false MALDI-TOF data in the PNAS 2016 paper. The Committee finds that while there were obvious errors published, [REDACTED] do not constitute an intentional act of Research Misconduct (Falsification) by Dr. Zhang, as described in the Policy IIIA and 45 C.F.R. § 689.2(c)(2).

By clear and convincing evidence, the Committee finds by a vote of 0 in favor and 3 against, that the Respondent, Dr. Zhang intentionally, knowingly or recklessly reported false MALDI-TOF data in the PNAS 2016 paper. The Committee finds that while there were obvious errors published, [REDACTED] do not constitute an intentional act of Research Misconduct (Falsification) by Dr. Zhang, as described in the Policy IIIA and 45 C.F.R. § 689.2(c)(2).

### **Summary of Investigation Committee Conclusions**

This has been a very difficult case for the Committee. The evidence clearly points to a situation where [REDACTED] Dr. Zhang did not have the requisite scientific domain expertise to provide the appropriate oversight or ensure the scientific rigor necessary to guarantee that the data being generated and ultimately reported were accurate or valid.

While Dr. Zhang has consistently stated that he has no expertise in molecular biology [REDACTED] the majority of the Committee believes that it was Dr. Zhang's responsibility to ensure the validity of the results. This was especially true once they started receiving reviews with specific concerns regarding the origin and identity of the IAGP sequence. The majority of the Committee cannot overlook the number of intentional misrepresentations in the manuscript, the lack of experimental rigor



with which the experiments were conducted, and the deliberate indifference demonstrated by Dr. Zhang regarding the integrity of the scientific record, especially in light of the very troubling concerns raised by reviewers. [REDACTED] corresponding author of the work, it is Dr. Zhang's responsibility to ensure the validity and accuracy of the reported results, and this was found to be woefully inadequate in this situation.

Regarding Dr. Zhang's other work, the committee did not analyze it in detail, because the nature of the allegations in this case were topic-specific and did not apply to Dr. Zhang's work in general. Specifically, the committee notes that Dr. Zhang's other work does not include a specific DNA analysis used to identify the Ivy arabinogalactan protein (IAGP). The nature of the misrepresentation of the data in Allegation #1 was a combination of a misrepresentation of how the data were acquired and how they were analyzed, in particular the fact that the comparisons were deliberately limited to only Arabidopsis. Second, the allegations specifically involve [REDACTED]

The committee finds that the Research Misconduct did not result in any specific harm to research subjects, researchers other than Dr. Zhang [REDACTED], or the public beyond what is noted in this report. The Ohio State University now mandates training in the responsible conduct of research for all researchers, and the University's Research Compliance/Research Integrity Officer notes that Dr. Zhang has not yet completed this requirement.

In Summary, the Committee finds, at both the preponderance of evidence and clear and convincing standards, that Dr. Zhang did commit Research Misconduct by the Reckless Falsification of the reported origin and sequence of the *IAGP* gene in PNAS 2016. The Committee finds that by a preponderance of evidence standard, that Dr. Zhang did not commit Research Misconduct regarding the reported location or methods reported regarding the MALDI-TOF data in PNAS 2016.

### **Consideration of Respondent's Comments**

On May 22, 2018, Dr. Zhang was provided the COEIC's Preliminary Investigation Report and all referenced documents and given thirty (30) days in which to review and provide comments to the Preliminary Report in accordance with Policy Section IV.F.5.b. On June 25, 2018, the university received comments submitted on Dr. Zhang's behalf by Mr. Paul Thaler<sup>83</sup>. On June 26, 2018, Dr. Zhang confirmed to Dr. Yucel via email that those were the comments that Dr. Zhang wanted the COEIC to consider<sup>84</sup>. The COEIC carefully reviewed Dr. Zhang's comments and determined that Dr. Zhang did not dispute the finding that the data in question under allegation #1 were falsified; that while Dr. Zhang did not agree with the finding of Research Misconduct, Dr. Zhang did agree to accept the proposed sanctions; and that Dr. Zhang did not provide any new facts or evidence in Dr. Zhang's comments.

Based on all of the evidence reviewed, including Dr. Zhang's comments to the Preliminary Report, the COEIC affirms its earlier determination by majority vote, at both the preponderance of the evidence, and clear and convincing standards, that Dr. Zhang acted recklessly by reporting falsified sequence data alleged to represent Ivy AGP, and by reporting a misrepresentation of the origin of those data in the PNAS paper. We affirm that Dr. Zhang's action(s) do constitute Research Misconduct (Falsification), as described in the Policy IIIA and 45 C.F.R. § 689.2(c)(2). A full description of the COEIC's consideration of Dr. Zhang's comments is provided separately in a letter to Dr. Zhang from Dr. Joseph Heremans, Chair, College of Engineering Investigation Committee, dated July 31, 2018<sup>85</sup>.

<sup>83</sup> ATT-55 - 2018-06-25 MingjunZhang's Comments to Preliminary Investigation Report\_Final

<sup>84</sup> ATT-56 - 20180626 - Email Zhang to RIO - confirming version

<sup>85</sup> ATT-57 - 20180731 - Letter Chair to Respondent - Consideration of comments to PR



### **Corrections/Retractions of the Scientific Record**

Per the Policy, section V.G, in cases where there is a finding of Research Misconduct, the University is required to seek to correct the published record with the assistance of the Respondent (Dr. Zhang), to the extent possible. Given the extent of the incorrect information and the finding of Reckless Research Misconduct by Dr. Zhang, the COEIC recommends that the University require the immediate retraction of the PNAS 2016 paper. In addition, the COEIC also recommends that the GenBank entries (KF752597, KM820289 and AKN58855) for *IAGP* also be pulled from the national database. Dr. Zhang should assist the University in ensuring that both of these actions take place immediately upon finalization of this report.

### **Recommended Sanctions**

Under the University Policy and Procedures Concerning Research Misconduct, section IV.F.5, the Committee shall include recommended sanctions in cases where allegations of Research Misconduct are substantiated. Further, under section IV.F.6, "Sanctions shall be commensurate with the severity of the research misconduct."

Given the specifics of this case, the Committee recommends the following sanctions:

1. The College of Engineering and Graduate School require that there be a co-advisor appointed for all students overseen by Dr. Zhang for a period of three (3) years from the date of the final report. For any current students, the College and Graduate School should review and determine if a co-advisor is needed based on the seniority/status of the student.
2. Dr. Zhang complete the CITI Responsible Conduct of Research training within two (2) weeks from the date of the final report.
3. Dr. Zhang be required by the College of Engineering to have experienced, qualified domain-specific co-Principal Investigators on all multi-disciplinary grant proposals that include research in areas of science outside Dr. Zhang's specific expertise for a period of three (3) years from the date of the final report.

### **Length of Proceedings**

This has been a very complicated, confusing and technically difficult case, with the majority of the relevant parties no longer at OSU, or no longer located in the US. In addition to the difficulty with accessing the various involved parties over the course of the investigation, [REDACTED]

The Committee has taken all appropriate steps to conduct this matter as quickly as possible while also ensuring the appropriate due diligence. All necessary extension requests were obtained by the University during this period.





## Appendix

**Complainant:** Anonymous

**Respondent:** Dr. Mingjun Zhang, Professor, Department of Biomedical Engineering

**Respondent Advisor:** Dr. Keith Gooch, Professor, Department of Biomedical Engineering

### Known PHS Federal Support

- Army Research Office (ARO) W911NF-10-1-0114 and ARO W911NF-12-1-0294;
- National Science Foundation (NSF) Civil, Mechanical, and Manufacturing Innovation #1029953,
- NSF Chemical Bioengineering, Environmental and Transport Systems # 0965877;
- Department of Energy (DOE) Bioenergy Science Center #DE-AC05-00OR22725.

### OSU Office of Legal Affairs Staff

- Emily Q. Schriver, Associate General Counsel, Office of Legal Affairs

### Correspondence and Attachments

ATT-1 - University Policy and Procedures Concerning Research Misconduct

ATT-2 - PNAS 113 23 E3193-3202 + SI

ATT-3 - 20160825 - Preliminary Assessment [REDACTED]

ATT-4 - 20160823 - Transcript Notification of Allegations Zhang #1

ATT-5 - 20160823 - Transcript RIO meeting [REDACTED] Zhang #2

ATT-7 - 20170321 - CII Preliminary Report – [REDACTED]

ATT-8 - CII A27 - Respondents response to PR (folder)

ATT-9 - 20170428 - CII Final Report – Zhang [REDACTED]

ATT-10 - CII A34 - 20170428 - Letter Chair to Respondents – Reconsideration

ATT-11 - CII A36 - 20170519 - SVPR to Respondents and Chair – Appeal Response

ATT-12 - CII A37 - 20170525- Email SVPR to RIO – approval Deadline extension and approval for external notification - Zhang case

ATT-13 - 20170721 - CII Revised Final Report – Zhang [REDACTED]

ATT-14 - 20170913 - COEIC Charge - Zhang

ATT-15 - 20170913 - Investigation Referral Letter for Zhang I-17-0092-O

ATT-16 - CII A14 - Nature Plants Materials (folder) and Data provided to COEIC 20171109

ATT-17 - 2017114 - Data provided to COEIC (folder)

ATT-18 - 20171128 - Data provided to COEIC (folder)

ATT-19 - 20171130 - Data provided to COEIC (folder)

ATT-21 - Audio recordings August 2016 (folder)

ATT-22 - 20180118 - Data provided to COEIC (folder)

ATT-23 - 20180122 - Data provided to COEIC (folder)

ATT-24 - 20180130 - Data provided to COEIC (folder)

ATT-25 - 20180301 - Response from Li Tan (folder)

ATT-26 - Emails July 3, 2015- Fwd\_Decision on manuscript NCOMMS-14-00999A-Z



[REDACTED]  
ATT-28 - 20180123 - COEIC Interview Zhang #1 and exhibits (folder)

[REDACTED]  
ATT-30 - 20180323 - COEIC Interview Zhang #2 and exhibits (folder)

ATT-31 - 20180328 - Documents from Zhang (folder)

ATT-32 - 20171212 - Email RIO to Respondents – Notification of external contacts

ATT-33 - 20170124 - Letter RIO to Carpenter – 1KP

ATT-34 - 20171212 - Letter RIO to Wang

ATT-35 - 20171212 - Letter RIOs to Mclab

ATT-36 - 20171210 - SVPR approval for external contacts

ATT-37 - 20180226 - RIO to Zhang [REDACTED] Notify external contact Tan

ATT-38 - 20180222 - ISVPR Approval for external notification

ATT-39 - 20180226 - Letter RIO to Tan – Investigation committee questions and attachment

ATT-40 - 20160523 - Zhang Press Release – PNAS paper

[REDACTED]  
ATT-42 - CII A42 - 31214 IVY-NP

ATT-43 - CII A43 - UTK IVP NP #1 mass spec trace

ATT-44 - 20170913 - Email NSF to RIO – Zhang Letter

ATT-45 - 20170918 - Email Mankowski to NSF – Request OSU policies

ATT-46 - 20170921 - Email Behnfeldt to NSF - Clarification

ATT-47 - 20171009 - Email Mankowski to NSF – Requested CVs

ATT-48 - 20171013 - Email NSF to Yucel – Referral letter

ATT-49 - 20171019 - Email Yucel to NSF – RE Referral letter

ATT-50 - 20171109 - Email Mankowski to NSF – Consultant CV

ATT-51 - 20180313 - Letter Yucel to NSF – Extension Request

ATT-52 - 20180313 - Email NSF to Yucel - NSF granting extension request

ATT-53 - 20180426 - Letter Yucel to NSF

ATT-54 - NSF OIG Assessing Intent in Research Misconduct Investigations

ATT-55 - 2018-06-25 MingjunZhang's Comments to Preliminary Investigation Report\_Final

ATT-56 - 20180626 - Email Zhang to RIO - confirming version

ATT-57 - 20180731 - Letter Chair to Respondent - Consideration of comments to PR

ATT-58 - 20180611 - Letter RIO to Runko - Status of OSU Investigation Redacted

ATT-59 - 20180612 - Email Runko NSF to RIO - extension granted and PR request\_Redacted