DNAku Consumers Profile: One of The First Direct to Customer Genetics Testing in Indonesia

Deby Erina Parung¹, Kians Azizatikarna¹, Dian Amirulloh¹, Erlin Listiyaningsih¹, Bharuno Mahesworo², Arif Budiarto²³, Simon¹, Bens Pardamean²⁴

¹Genetics Indonesia, Jakarta, Indonesia 12940
²Bioinformatics & Data Science Research Center Bina Nusantara University Jakarta, Indonesia 11480
³Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
⁴Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia 11480

bharuno.mahesworo@binus.edu

Abstract. DNAku test is one of the first direct to customer genetics testing in Indonesia. DNAku test came with a different approach as it gave their customer reports of their genetic predisposition to corresponding traits. Four months after launch, we create consumer report of this genetics testing based on DNAku questionnaire results. We applied three statistical methods, including descriptive statistics, multivariate normality test, and t-test to the demographic data of the DNAku customers. The result shows population clusters which may represent population groups that are interested in genetics testing. The questionnaire also has questions about the customer’s lifestyle. The results of these questions were processed to find the lifestyle of DNAku customers.

Keywords—Genetics Testing, Consumers Profiling, Healthy Lifestyle, Genetics Profile

1. Introduction
The scientific advancement and new findings in genetics have helped us in many ways such as finding the identity of a deceased body or determining whether an individual is the biological parent of another individual. Genetic testing also has the possibility of provides individuals with a list of diseases that the individuals may develop in the future. However, with our current understanding of genetics, the results of this kind of prediction are still in percentage and the reliability of the result is questionable. In developed countries such as the United States, European countries, Japan and South Korea, genetic testing has become more common, compared to the less developed countries or developing countries such as Indonesia due to the expensive investment of genetics testing infrastructure, technologies and science.

In the United States, people are interested in genetics testing because of different reason. The American native including Inuit or Indian is now a minority in the United States [1]. The majority who
live in the United States is whether Caucasian from Europe, African, East Asian or South Asian immigrant descendants. Their ancestors may have come to North America for a long time, and they may have children with a person from different ethnicity. This situation resulted in the mixing of racial genes in the offspring of those currently living in the United States [2]. Hence, people want to find where they are actually coming from. A direct to customer (DTC) genetics testing service in the United States, 23andMe [3], offer a genetics ancestry testing to discover where in the world your DNA is from. This kind of test is very popular in the United States and became one of the reasons why genetics testing became very popular lately.

Based on the authors' opinion, this case may not suitable in the Indonesia market, since most of the Indonesians know what ethnicity they are or simply do not care. In fact, the human migration path also showed that South-East Asian especially Indonesian are coming from the same ancestor. Other ethnicity groups such as Chinese-Han, Arabs and South Asians are also present in the Indonesian population as minorities. Therefore, the curiosity of Indonesians to find their ancestry group is possibly low.

Genetic Indonesia, one of the first DTC genetics testing company, comes with a distinct approach. They offer a genetics-test that will inform the customer with a kind of lifestyle that suits you the most. The test is branded with the name of “DNaku” [4] which means “my DNA” in English. The test was introduced on 17th August 2019 and after 4 months, it has more than 400 processed samples. In this paper, we analyzed the demographic of the first 259 samples tested in DNAku. The data studied in this research were anonymous and they completed a brief informed consent form for this observational study. The result of this paper shows the demographics of the participant which represent the demographic group interested in genomics testing.

In the last decade, health practices have relied on the analysis of genomics data to support the prognosis and diagnosis due to reduced cost of this data capturing process through the next-generation sequencing and microarray genotyping [5–8]. This data has also been the primary study object in various research related to certain diseases, traits, or ancestry studies [9–14]. In the beginning, genetics testing was done very limited merely on a clinic or hospital setting. However, the emergence of disruptive information technology has driven some changes in these practices [15,16]. As can be found in financial sectors, where commerce and banking practices have shifted to a digital environment [17,18], health practices have also started to evolve [19,20]. It is indicated by the rise of health-tech start-ups globally.

In the genetics testing sector, a new way of providing this service by relying on the digital transaction and process was initiated back in 2003 as reported in a study by Williams-Jones [21]. Even in 2001, a study has forecasted that in the 2010s the digital practices for genetic testing would be ubiquitous [22]. This innovative practice is intended to provide a more personal genomics service for the consumers. To date, there are more than 20 companies still actively advertise their DTC genetics testing services which mainly based in the US and some others in Europe, Asia, and Australia [23].

In Indonesia, this kind of service is still considered rare. DNAku is the pioneer of this service in Indonesia. DNAku utilized modern genotyping technology and sophisticated bioinformatics tools in analyzing genetics data. Together with individuals' environment and lifestyle, genetics has an important role in individual’s life [24,25]. These factors could help the researcher in Indonesia develop their health monitoring system which is being researched for the last few years [26–29]. The data collected by DNAku also can be used to strengthen the cancer early detection system [30–34] or cardiovascular diseases early detection system [35,36]. Many genetic studies findings explain gene expressions and significant markers relevant to certain traits. Those traits could be phenotype appearance, including eyes colour, skin colour, etc.; or risk of getting certain diseases or other expressions such as lactose intolerant. DNAku test matches its customers' DNA with these findings. Currently, there are four modules of the test that can be bought separately: skin, sport, diet and health. Each module gave the customer a report that shows customers' genetic predisposition or in other words, characteristics, to corresponding traits that related to the module.

The test started with epithelial tissue collection using a buccal swab. The customer could do it by themself since the buccal swab test kit could be sent to customers' home and the sample could also be posted back to DNAku lab. The collected tissue is then processed to extract the customer's DNA. To get
the genetic profile, a genotyping test is applied to the DNA. DNAku laboratory is equipped with Illumina Microarray to detect more than 660,000 target markers. For the Microarray, DNAku specifically uses Illumina Asian Screening Array Bead Chip, which includes significant markers, based on studies on Asian Population. The genotyping result is then analyzed using a bioinformatics approach to create the customer's characteristics.

2. Materials and Method
The data was processed and analyzed using the Python programming language. The demographic of the sample shown in descriptive statistic parameters, including mean, standard deviation, the minimum and maximum value of each variable. A descriptive statistic is used to describe the central tendency and the dispersion of quantitative data [37]. A histogram plot is also created using the seaborn library to help authors decide whether certain variables are normally distributed or not. A Kernel Density Estimate (KDE) plot was also produced to visualizing the probability density of continuous variables [38]. KDE plot helped the author on determining the characteristic of the samples based on two closely related variables. Lastly, t-test was applied to some variables to find a significant difference between groups [39]. The method in this study described in Figure 1.

![Figure 1. Research Workflow](image)

2.1. Questionnaire
The questionnaire asked in DNAku test includes basic demographic, lifestyle, hereditary disease history and hereditary disease family history. The questions were asked to the sample via DNAku website, and the samples answer the questionnaire based on self-claim. The data obtained is stored in DNAku along with the user’s consent. In this paper, we focused on demographic and lifestyle questions. The questions can be found in Table 1.

| No | Question | Answer |
|----|----------|--------|
| 1  | Gender   | Radio  |
|    |          | 1. Male |
|    |          | 2. Female |
| 2  | Weight   | Integer |
| 3  | Height   | Integer |
| 4  | Date of Birth | Date |
| 9  | What ethnicity are you? (Can be more than 1) | Checklist |
|    |          | 1. Ambon |
|    |          | 2. Balinese |
|    |          | 3. Banjar |
|    |          | 4. Banten |
|    |          | 14. Malay |
|    |          | 15. Manado |
|    |          | 16. Minangkabau |
|    |          | 17. Papuan |
|   | Languages                 |
|---|--------------------------|
| 5. | Batak                    |
| 6. | Betawi                   |
| 7. | Bugis                    |
| 8. | Cirebon                  |
| 9. | Dayak                    |
| 10. | Flores                  |
| 11. | Javanese                 |
| 12. | Madurese                 |
| 13. | Makassar                 |
| 18. | Sumbanese                |
| 19. | Sundanese                |
| 20. | Arab                     |
| 21. | Caucasian                |
| 22. | Chinese                  |
| 23. | Indian                   |
| 24. | Crossbreed               |
| 25. | Other (Specify)          |

10. How often do you eat red meat? Radio

1. I don’t eat red meat
2. 1-2 times per week
3. 3-4 times per week
4. More than 4 times a week

11. How often do you eat vegetables? Radio

1. I don’t eat vegetable
2. 1-2 times per week
3. 3-4 times per week
4. More than 4 times a week

12. How often do you do exercise? Radio

1. I never exercise
2. 1-2 times per week
3. 3-4 times per week
4. More than 4 times a week

13. Are you a smoker? Radio

1. Yes, but I have stopped
2. Yes, I smoke sometimes
3. Yes, I smoke every day
4. No

16. How often do you drink alcohol? Radio

1. More than three times a month
2. 2-3 times per month
3. Once a month or less
4. Never

3. Results and Discussion

After processing the DNAku data using a descriptive statistical method, the description of the sample demographic can be found in Table 2. The data is also plotted in histogram charts, Figure 2 and Figure 3, where the samples were divided into gender group and plotted according to its age and BMI distribution.

| Gender | N (%) | Mean | Stdev | Min - Max |
|--------|-------|------|-------|-----------|
| Male   | 122 (47.10) | - | - | - |
| Female | 137 (52.90) | - | - | - |

| Age (years) | N (%) | Mean | Stdev | Min - Max |
|-------------|-------|------|-------|-----------|
| All         | - | 36.36 | 11.77 | 7.94 – 87.07 |
| Male        | - | 39.22 | 12.68 | 12.95 – 87.07 |
| Female      | - | 33.82 | 10.29 | 7.94 – 70.98 |
### Weight (kg)

|        | All   | Male  | Female |
|--------|-------|-------|--------|
|        | 66.68 | 74.01 | 60.16  |
|        | 14.44 | 12.16 | 13.16  |
|        | 30.0 – 110.0 | 48.0 – 110.0 | 30.0 – 100.0 |

### Height (cm)

|        | All   | Male  | Female |
|--------|-------|-------|--------|
|        | 164.30| 170.46| 158.81 |
|        | 8.76  | 6.98  | 6.14   |
|        | 130 – 190 | 150 – 190 | 130 – 175 |

### BMI (kg/m²)

|        | All   | Male  | Female |
|--------|-------|-------|--------|
|        | 24.59 | 25.45 | 23.83  |
|        | 4.47  | 3.74  | 4.91   |
|        | 15.57 – 38.09 | 16.42 – 38.09 | 15.57 – 38.09 |

### Ethnicity

| Ethnicity | All (No. %) |
|-----------|-------------|
| Chinese   | 112 (42.11) |
| Javanese  | 76 (28.57)  |
| Sundanese | 31 (11.65)  |
| Betawi    | 17 (6.39)   |
| Malay     | 17 (6.39)   |
| Batak     | 16 (6.01)   |
| Minang kabau | 8 (3.01) |
| Madurese  | 7 (2.63)    |
| Bugis     | 6 (2.26)    |
| Arab      | 4 (1.50)    |

---

**Figure 2.** Age Distribution Among Gender Group (p-value = 0.00019)
The sample demographic appears to be almost balance between gender group with 122 for male samples and 137 for the female sample. With a total of 259 samples, the male to female sample proportion is 47.1 to 52.9. According to Figure 2, the age distribution for male sample peaks at ± 30 years old and ± 50 years old while female sample age distribution peaks at ± 25 years old. Both of them show normal distribution with a small skew to the right. Even though the ratio between male and female is almost equal, the age distributions appear very different. The t-test between gender groups on their age resulting in a significant value, 0.00019.

On the weight and height variables, the male samples are heavier and taller as the mean on those variables shown. The BMI mean also shows that the female samples are smaller. Based on histogram plots in Figure 3, BMI distribution also normally distributed among gender groups. According to the t-test result (p-value = 0.00327), BMI variable is also a significant variable for differentiating sample by gender. This phenomenon is hard to compare to the global population since at the time this paper is written, the authors could not find BMI distribution research in the Asian population, especially the Indonesian population. Moreover, the tendency of becoming overweight/obese between gender groups is different among races [40,41] and age group [42]. Therefore, we cannot compare this phenomenon to research findings that used different population race or that used mixed-race samples.
From the age and BMI multivariate distribution among gender group in Figure 4, the result indicates that people who interested in this kind of test are whether young females with normal BMI classification or males aged 30 – 50 years old with higher BMI. The reasons of this phenomenon need to be identified further and might be more relevant using social science approaches.

The ethnicity question in DNAku test will be used for future study related to ethnicity. Right now, the data from this question can be used to identify which ethnicity groups that are interested in this kind of test. Surprisingly, most of the samples are Chinese (42.11%), followed by the Javanese and Sundanese, 28.57% and 11.65% respectively. Based on Indonesia’s website, Chinese ethnicity is a minority in Indonesia with only 1.2% of the whole population. Meanwhile, Javanese and Sundanese are 40.22% and 15.5% respectively [43].

Table 3. Samples Lifestyle

| Question      | Never | 1-2x per week | 3-4x per week | >4x per week |
|---------------|-------|---------------|---------------|--------------|
| Red Meat      | 15    | 175           | 54            | 22           |
| Vegetable     | 9     | 77            | 100           | 80           |
| Exercise      | 58    | 139           | 48            | 21           |
| Smoking       | 215   | 24            | 18            | 9            |
| Alcohol       | 194   | 48            | 17            | 7            |

Besides the basic demographic information, DNAku also captures lifestyle information which includes meat and vegetable consumption, exercise, smoking, and alcohol habit. The summary of lifestyle information from the samples is shown in Table 3. Majority of the samples consume more vegetable than red meat. Given that most of the samples are young productive people who live in crowded cities like Jakarta, their exercise habit indicating a relatively good lifestyle with the majority of them do physical activities 1-2 times per week. Additionally, this group of samples also adopt a healthy lifestyle where 70% of them avoiding smoking and alcohol consumption, which scientifically significant as causes of several diseases[44,45].

4. Conclusion

The DNAku data reveal that this kind of test is attractive for both male and female since the demographic show almost equal proportion between gender group. The female samples are mostly young adult with normal BMI category. Meanwhile, the male samples are mostly aged between 30 and 50 years old and have higher BMI compared to female samples. The samples also mostly came from Chinese ethnicity which is a minority in the Indonesian population. Also, most of the samples are having a healthy lifestyle, including eating less meat, eat more vegetable, do not smoking and do not drink alcohol.

References
[1] Grossman DC, Krieger JW, Sugarman JR, Forquera RA. Health Status of Urban American Indians and Alaska Natives: A Population-Based Study. JAMA 1994;271:845–50. https://doi.org/10.1001/jama.1994.03510350055037.
[2] Emeka A, Vallejo JA. Non-Hispanics with Latin American ancestry: Assimilation, race, and identity among Latin American descendants in the US. Soc Sci Res 2011;40:1547–63. https://doi.org/https://doi.org/10.1016/j.ssresearch.2011.06.002.
[3] 23andMe n.d. https://www.23andme.com/.
[4] Genetic Indonesia. DNAku 2017.
[5] Yusuf I, Miskad UA, Lusikooy RE, Arsyad A, Irwan A, Mathew G, et al. Genetic Risk Factors for Colorectal Cancer in Multiethnic Indonesians. BioRxiv 2019. https://doi.org/10.1101/626739.
[6] Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 2012;8. https://doi.org/10.1371/journal.pcbi.1002822.

[7] Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. Stat Med 2015;34:3769–92. https://doi.org/10.1002/sim.6605.

[8] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017. https://doi.org/10.1016/j.ajhg.2017.06.005.

[9] Budiarto A, Mahesworo B, Baurley J, Suparyanto T, Pardamean B. Fast and Effective Clustering Method for Ancestry Estimation. Procedia Comput Sci 2019;157:306–12. https://doi.org/10.1016/j.procs.2019.08.171.

[10] Zhang B, Jia WH, Matsuo K, Shin A, Xiang YB, Matsuda K, et al. Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. Int J Cancer 2014;135:948–55. https://doi.org/10.1002/ijc.28733.

[11] Mcmahon C, Baurley J, Bridges W, Joyner C. A two-phase Bayesian methodology for the analysis of binary phenotypes in genome-wide association studies. Biometrical J 2020;62:191–201. https://doi.org/10.1002/bimj.201900050.

[12] Lumbanraja FR, Mahesworo B, Benggorro TW, Baurley J, Pardamean B. An Evaluation of Deep Neural Network Performance on Limited Protein Phosphorylation Site Prediction Data. Procedia Comput Sci 2019;157:25–30. https://doi.org/10.1016/j.procs.2019.08.137.

[13] Roberts JS, Ostergren J. Direct-to-Consumer Genetic Testing and Personal Genomics Services: A Review of Recent Empirical Studies 2013:182–200. https://doi.org/10.1007/s40142-013-0018-2.

[14] Gai K, Qiu M, Applications XS-J of N and C, 2018 undefined. A survey on FinTech. Elsevier n.d.

[15] Lee I, Horizons YS-B, 2018 undefined. Fintech: Ecosystem, business models, investment decisions, and challenges. Elsevier n.d.

[16] Hwang J, Christensen CM. Disruptive innovation in health care delivery: A framework for business-model innovation. Health Aff 2008;27:1329–35. https://doi.org/10.1377/hlthaff.27.5.1329.

[17] Williams-Jones B. Where There’s a Web, There’s a Way: Commercial Genetic Testing and the Internet 2003:46–57. https://doi.org/10.1159/000069538.

[18] Collins FS, Mekusick VA. Implications of the Human Genome Project for Medical Science 2001;285.

[19] Avinun R. The E Is in the G: Gene–Environment–Trait Correlations and Findings From Genome-Wide Association Studies. Perspect Psychol Sci 2019;15:81–9.
[26] Kacamarga MF, Budiarto A, Pardamean B. A Platform for Electronic Health Record Sharing in Environments with Scarce Resource Using Cloud Computing A Platform for Electronic Health Record Sharing in Environments with Scarce Resource Using Cloud Computing. Int J Online Biomed Eng 2020;16:63–76. https://doi.org/10.3991/ijoe.v16i09.13187.

[27] Zulkifli R, Pardamean B. Design of A Drug Testing Online Portal. Procedia Eng 2012;50:38–44. https://doi.org/10.1016/j.proeng.2012.10.005.

[28] Pardamean B, Anindito, Djoenang A, Tobing N. Disease Management Information System. Am J Appl Sci 2013;10:724–33. https://doi.org/10.3844/ajassp.2013.724.733.

[29] Pardamean B, Wibisono KA. Nutrition Management and Diet Monitoring Information System. Res J Appl Sci 2014;9:412–7. https://doi.org/10.3923/rjasci.2014.412.417.

[30] Pardamean B, Suparyanto T. Hospital-based cancer registry application. 2017 Int. Conf. Inf. Manag. Technol., 2017, p. 44–8. https://doi.org/10.1109/ICIMTech.2017.8273509.

[31] Muljo HH, Pardamean B, Perbangsa AS. The Implementation of Online Learning for Early Detection of Cervical Cancer The Implementation of Online Learning for Early Detection of Cervical Cancer. J Comput Sci 2017;13:600–7. https://doi.org/10.3844/jcssp.2017.600.607.

[32] Muljo HH, Perbangsa AS, Pardamean B. Mobile Learning for Early Detection Cancer. Int J Interact Mob Technol 2018;12:39–53.

[33] Muljo HH, Perbangsa AS, Pardamean B. Improving Early Cancer Detection Knowledge through Mobile Learning Application. Int J Online Biomed Eng 2019;15:60–70.

[34] Muljo HH, Pardamean B, Perbangsa AS, Purwandari K, Mahesworo B, Hidayat AA, et al. TAM as a Model to Understand the Intention of Using a Mobile-Based Cancer Early Detection Learning Application. Int J Online Biomed Eng 2020;16:80–93.

[35] Subagyo A, Pardamean B, Abbas BS. Expert-System Based Medical Stroke Prevention. J Comput Sci 2013;9:1099–105. https://doi.org/10.3844/jcssp.2013.1099.1105.

[36] Caraka RE, Nugroho NT, Tai S. Feature importance of the aortic anatomy on endovascular aneurysm repair (EVAR) using Boruta and Bayesian MCMC. Commun Math Biol Neurosci 2020;2020.

[37] Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 2019;22:67–72. https://doi.org/10.4103/aca.ACA_157_18.

[38] Rudemo M. Empirical Choice of Histograms and Kernel Density Estimators. Scand J Stat 1982;9:65–78.

[39] Kim TK. T test as a parametric statistic. Korean J Anesthesiol 2015;68:540–6. https://doi.org/10.4097/kjae.2015.68.6.540.

[40] Paeratakul S, White MA, Williamson DA, Ryan DH, Bray GA. Sex, Race/Ethnicity, Socioeconomic Status, and BMI in Relation to Self-Perception of Overweight. Obes Res 2002;10:345–50. https://doi.org/10.1038/oby.2002.48.

[41] Camhi SM, Bray GA, Bouchard C, Greenway FL, Johnson WD, Newton RL, et al. The Relationship of Waist Circumference and BMI to Visceral, Subcutaneous, and Total Body Fat: Sex and Race Differences. Obesity 2011;19:402–8. https://doi.org/10.1038/oby.2010.248.

[42] Evans EM, Rowe DA, Racette SB, Ross KM, McAuley E. Is the current BMI obesity classification appropriate for black and white postmenopausal women? Int J Obes 2006;30:837–43. https://doi.org/10.1038/sj.ijo.0803208.

[43] Indonesia go id. Suku Bangsa 2017.

[44] West R. Tobacco smoking: Health impact, prevalence, correlates and interventions. Psychol Health 2017;32:1018–36. https://doi.org/10.1080/08870446.2017.1325890.

[45] Aboulmaouahib S, Madkour A, Kaarouch I, Sefrioui O, Saadani B, Copin H, et al. Impact of alcohol and cigarette smoking consumption in male fertility potential: Looks at lipid peroxidation, enzymatic antioxidant activities and sperm DNA damage. Andrologia 2018;50:e12926. https://doi.org/10.1111/and.12926.