

Know thyself: Data driven self-awareness for understanding our unconscious behaviors

-Citizen Science Genetics for sleep and wellness research-

Takashi Kido^{a,b} & Melanie Swan^c

^aRiken Genesis Co., Ltd., Taito-ku, Taito, 1-5-1, Tokyo 110-8560, Japan

^bJST PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

^cMS Futures Group, P.O. Box 61258, Palo Alto, CA 94306, USA

Correspondence Email: kido.takashi@gmail.com

Abstract

We have been testing the hypotheses that our unconscious behaviors are often more predominant than our conscious behaviors. We have tried to quantify the degree of influence of our genetic factors and unconscious signals from the environment in our behaviors. In this paper, the concept and the research framework of MyFinder proposed by the authors are discussed. MyFinder combines the notion of intelligent agent from the artificial intelligence (AI) community and personal genome research results from the biomedical community. One of the important objectives of MyFinder is to allow the person to be conscious of (i.e., be aware of) her/his unconscious behaviors. The on-going experiments related to the Quantified Self (QS) and Citizen Science, as well as author's real experiences are introduced. We discuss the meaning of self-awareness and the social impact of our findings.

Introduction

Socrates once said, "know thyself." A desire to know more about myself was the springboard for my concept. This desire can lead one to ponder the philosophical question, "who am I?" We take for granted the fact that we actively make life-orientating decisions using our own free will. However, what if these decisions are in fact controlled by our genetic disposition and unconscious signals from our societal surroundings?

The arrival of the Personal Genome Age has resulted in the current boom in the field of gene analysis. The concept of Personal Genome refers to the gene information about a single person. Individual gene differences are associated with such factors as susceptibilities to various disorder [PG 13]s, physical constitution [Ashley 10], and the effects of drugs [Butte 08]. Owing to revolutionary developments in gene analysis technology, we are approaching the age in

which each one of us can obtain our personal gene information [Jun W 08].

From an evolutionary prospective, it is likely that rapid modern changes in lifestyle largely influence the human mind and body. Our fundamental motivation for this research is our desire to enrich the minds of modern people by clarifying the mechanisms of the brain, genes, and body and correcting them through the use of artificial intelligence [Kido 10].

What kind of influence does personal genome information have on an individual's self-knowledge, our values, future lives, and research in artificial intelligence? This paper aims to outline the influence of personal genome information using the research framework of MyFinder.

MyFinder

MyFinder [Kido 11] has two objectives. The first objective is to find innate potential characteristics and personality and to bring out the maximum potential in one's abilities (individual approach). The second objective is to create a research platform that facilitates scientific discoveries through community computing (collective intelligence approach).

In terms of the first objective of MyFinder, recent Personal Genome research has focused on the realization of custom-made medical care by finding disease risks and drug effects, but MyFinder is unique as emphasis is placed on aspects relating to wellness, mental sciences, and behavioral sciences. MyFinder is based on recent findings relating to gene expression control and the epigenome [Sasaki 05] and supports the hypothesis that our physical,

chemical, and psychological stress greatly influences the activity of our genes. For instance, recent research has reported that the act of laughing affects diabetes [Hayashi 06], and positive mental stress caused by laughing is associated with turning on and off the genetic switch that controls gene expression at the cellular level. It will be possible for an intelligent agent to monitor our daily physical, chemical, and mental stress by way of everyday observation and analysis of our daily habits including eating, sleeping, work style, time management, social interaction, skills, and preferences. The intelligent agent technology can be effective in learning individual behavioral characteristics and stress status.

Regarding the other goal of MyFinder, to build a research platform for scientific discoveries through the use of community computing, MyFinder learns each user's personality by monitoring daily behavior and aims to interactively inform the user of its findings by using psychology-based and behavioral sciences-based findings (e.g. including the Enneagram theory in psychology [Helen 96]). This function will aid an individual in rediscovering his/her innate potentials and personality.

The Personal Genome information environment acts as an intermediary in the boundary region between artificial knowledge and the theory and practice of gene research. MyFinder is a new research framework which uses the Personal Genome information environment and integrates the concept of intelligent agent from artificial intelligence research and the concept of disease-related gene search in the community of genetics. This concept is shown in Fig. 1.

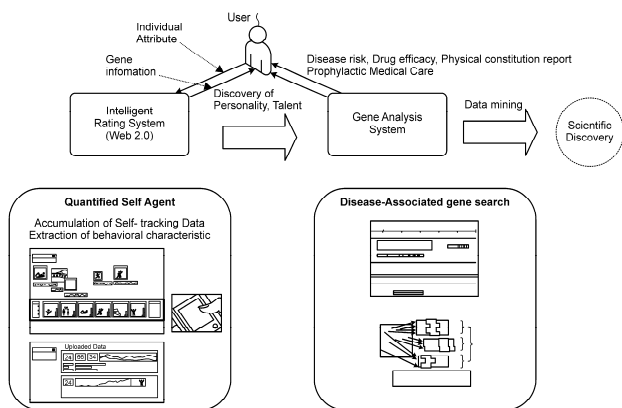


Figure 1. The Concept of MyFinder

Quantified Self and Exploration of the Mind and Body using Personal Genome

Quantified Self (QS) originated in Silicon Valley in 2011 and it is a movement that obtains new findings by

quantitatively observing human behaviors and status by using computers and various gadgets [QS 11]. We first encountered QS at the first Silicon Valley QS meeting and presented a paper on the Personal Genome experience at the San Francisco QS meeting in 2012. We started to take action in Japan in 2012 by launching QS Tokyo [QS 11].

We continue to work on recording awareness by recording daily lifelogs. Shown in Table 1 are examples of the lifelogs attempted (or considered) by the first author and supporters as an object of QS.

Table 1. Examples of Objects of QS

Measurement Category	Measurement Means	Measured Characteristics
Sleep	Accelerometer Actigraph	Sleeping time, quality and depth of sleep
Concentration/relax	Simplified electroencephalograph	α -wave, β -wave, θ -wave, strength, spectrum
Stress	Heart Rate Sensor	Stress intensity
Exercise	Accelerometer	Amount of exercise, calorie consumption
Meals	Camera	Contents and time of meals
Social interaction	SNS, Mail log	SNS activity
PC work environment	Mouse and typing operation monitoring	amount, time, recipient
Personality	Questionnaires	Work efficiency, interest
Physical constitution	Medical examination	Behavioral characteristics and tendency
Enteric activity	Feces	Various hormones, basal metabolism components
Daily event Emotion, feelings	Diary Diary, icon	Enteric bacterial activity
Cognitive characteristics	Questionnaires, cognition test	time, time of occurrence, category of emotion Various cognition tendencies

We regard both quantitative data (measurable data such as various sensor information, behavioral history, and answers to questionnaires) and qualitative data (text and images describing subjective awareness and feelings, which are difficult to measure quantitatively). One of the objectives of MyFinder is to allow the person to be conscious of (i.e., be aware of) her/his unconscious behaviors.

The ultimate objective of MyFinder is to identify the intelligent agent which watches over unconscious behaviors and state of a person and allows the person to increase his/her "awareness" by supporting the process of becoming conscious of his/her unconscious behavior. This means that the intelligent agent reads unconscious signals of the individual, finds a pattern from the signals,

constructs a model from the pattern, makes predictions based on the model, provides the person with predictions at the right timing, and continues to correct the model by receiving feedback in an interactive manner from the person. For instance, the revolutionary device such as the sociometer used in the research of “honest signals” by Pentland [Pentland 13] is an example of enabling the measurement of unconscious social signals of a person (nonverbal information such as a manner of speaking, tone of voice, and subtle actions).

Self-Discovery by Personal Genome

The Personal Genome has brought about innovations in the community of prophylactic medical care. Among recent prominent examples of progress in Personal Genome analysis, a state-of-the-art technology is the integrative Personal Omics Profile (iPOP) project. Rui et al. [Rui 12] integrated multi-omics data (biomolecular information including epigenome, transcriptome, proteome, metabolome, etc.) including the Personal Genome of Dr. Michael Snyder of Stanford University and information collected over a period of 38 months (as of May, 2013) and reported the occurrence of genome edits caused by infection with HRV or RSV. In the 14th month, the results of changes in various omics data, including gene information caused by an allergy-induced fluctuation in cytokine, were reported [Rui 12]. Dr. Snyder has revealed gene information relating to various diseases primarily on diabetes and has tried to make use of the information for disease prediction, early diagnosis, monitoring, and therapy.

The first author is proceeding with a gene analysis of himself in cooperation with the bioinformatics team of Stanford University. The left section of Fig. 2 is one example of the mapping of gene variation and predominant physical constitution, drug efficacy, behavioral characteristics, character traits (disease risk, virus/bacterium resistance, drug efficacy, senses, thoughts, behavioral characteristics, physical constitution, aging, and physical ability) of the human body. The right section of Fig. 2 shows the positions of the genes on human chromosomes that increase disease risks due to gene variations. Common Variants are ordinary variations for which the gene variation is 1% or more, and Rare Variants are rare variations for which the gene variation is less than 1%. A universal scientific standard of measure has yet to be developed for Personal Genome analysis, although various results are obtained such as disease risk prediction as pointed out by Venter et al. [Venter 10].

Furthermore, it is known that disease risk can be dependent on racial differences [Kamatani 10], [Kido 10]. The first author and co-researchers compared disease risk evaluations of the Personal Genome services of three companies and reported on the results, such as degree and cause of difference, comparisons among risk prediction algorithms, difference in disease risk between a Japanese

group and an American/European group. We are planning to conduct demonstrative research as well as research using mathematical models. As we pointed out in [Kido 06] and [Kido 11], there is ample room for adopting artificial intelligence technology for Personal Genome analysis.

The first author continues his attempts to annotate disease risk, drug efficacy and information concerning other genetic characterizations (e.g., reactions to alcohol, nature of hair, smoking habit, food preferences, life span, sensitivity to pain, and tendency to avoid failure) to better characterize his own genome information, as was partially introduced in [Kido 10]. For instance, he has reported research results based on themes such as heavy or light drinkers, smoker or non-smoker, long-lived or short-lived, gain weight easily or with difficulty, successful or unsuccessful with dieting, and tendency or not to avoid failure.

Examples of interpretation obtained from the Personal Genome of the first author and his father are introduced below.

- Difficulty in metabolizing alcohol (ALDH2)
- Special gene frame shift variation (TRPV1) reactive to capsaicin (pungent component of chili pepper).
- A gene variation (ACTN3) related to muscle performance that is the same as that of many world-class sprinters.
- Poor ability of avoid failure based on past experience (DRD2).
- High tendency to seek new experiences (DRD4).
- Plurality of disease candidates of higher risk than the average, although no serious disease risk can be found at present.
- The author’s father identified a risk for colorectal cancer from his Personal Genome, succeeded in early detection and excision of the cancer, and is currently very healthy.

Particularly evident with the last example of the author’s father is the fact that the Personal Genome can be applied in the future to prophylactic medical care.

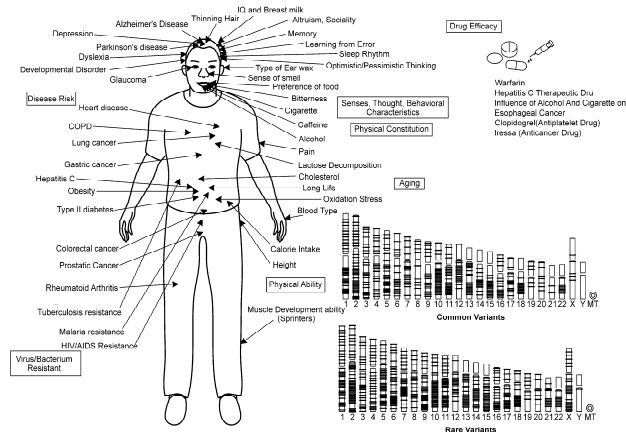


Figure 2. Self Discovery by Personal Genomes

The Personal Genome and related supporting sciences have the potential to discover a new approach in observing human beings, which will consequently lead to new perspectives on life. We would particularly like to note the possibility that mental stress exerts great influence on the activities of genes discovered through research on the epigenome in recent years [Sasaki 5]. We hope the study of individual differences (individuality) will lead to the enrichment of human nature, and not to discrimination. The aim of MyFinder is to help individuals to reconsider and positively understand themselves. In other words, it is intended to provide a device which enables us to experience the concept of “we are different and that is good” by identifying and accepting differences from others.

Scientific Discovery through the use of Community Computing: Citizen Science

Using the crowdsourced health study platform, Genomera, the Genetic Predisposition for Healthy Sleep (available at: <http://genomera.com/studies/sleep-genetic-predisposition-for-healthy-sleep>), a community computing peer study was launched in September 2012. As of January 2014, 26 participants had shared their 23andMe genomic data relevant to the study, and 12 had fully completed the supporting phenotypic questionnaire. The phenotypic questionnaire is the Pittsburgh Sleep Quality Index, a standard tool for measuring sleep quality in academic research studies. The aim of the study is to investigate possible linkage between personalized genomic profiles and sleep quality. The study was inspired by similar recent studies that linked genetic SNPs to sleeping profiles in healthy adults [Bodenmann 12, Nova 12, Landolt 11, Ciarleglio 8].

Demographically, the study had 12 participants (six males, four females, two undisclosed), nine of whom disclosed their age. The age range was 24-61 years, with three participants in their 20s, three in their 30s, two in their 50s, and one over 60. Participants primarily lived in a variety of cities in the United States, and a few lived internationally. Most participants were of European descent, and a few were of African and Asian descent.

The findings of the study appear in Figure 3. The phenotypic results are in the top line (“Instrument Score”), consisting of 15 questions which participants answered on a four-tier scale of bad-to-good. Given a maximum score of 45, these study participants indicated that they are already enjoying a fairly good quality of sleep (scores ranging from 21-42). The x-axis of Figure 3 represents the 12 participants with their phenotypic score and different genomic scores. The second line from the top indicates the total genomic score. This was calculated based on the total number of favorable alleles each participant had, based on the 11 SNPs (single nucleotide polymorphisms) examined in the study that are listed in Table 2. For each SNP, a

value of 0, 1, or 2 was assessed per the number of favorable alleles. A maximum value of 22 was possible. The study participants had an average score of 14 and a range of 9-18, indicating a higher than random or average presence of the favorable alleles. However, it can be seen readily from Figure 3 that no correlation was found between the phenotypic survey result and the genomic profile. Higher sleep survey scores were not indicative of a higher presence of the favorable alleles in the underlying genomic profile.

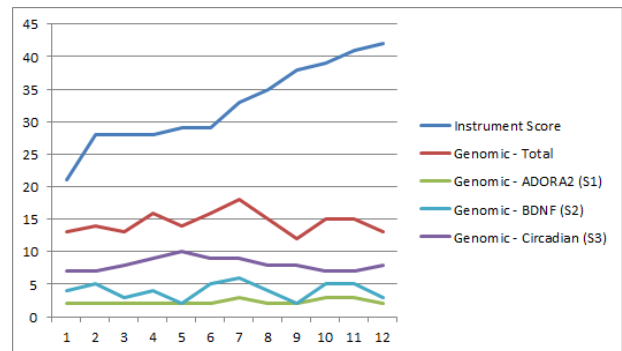


Figure 3. Genomics and Healthy Sleep Study Results

Three studies had found a correlation between genomic profile and sleep quality, and the lower three lines in Figure 3 enumerate them. First, the bottom line, looking at the two ADORA2 SNPs investigated by Bodenmann does show a slight incline in the right half of the line, but it is not striking enough to suggest a correlation. The second line from the bottom represents the BDNF, COM VALMET, and PRNP SNPs evaluated by Landolt, and shows no correlation with reported results from the phenotypic study. Likewise, the third line from the bottom represents the circadian rhythm SNPs delineated by Ciarleglio and shows no correlation with participant sleep quality.

Table 2. Genomic SNPs Examined

Gene	SNP	Favorable Allele	Reference
ADORA2A	rs5751876	T	Bodenmann 12
ADORA2A	rs2236624	C	Bodenmann 12
BDNF	rs6265	C	Landolt 11
COMT VALMET	rs4680	A	Landolt 11
PRNP	rs1799990	A	Landolt 11
ARNTL2	rs4964059	A	Ciarleglio 8
CLOCK	rs1801260	A	Ciarleglio 8
PER2	rs2304669	A	Ciarleglio 8
PER3	rs228697	C	Ciarleglio 8
AB047536	rs10462021	A	Ciarleglio 8
ADA	rs73598374	T	Nova 12

Overall, the citizen science study was unable to corroborate a correlation between genomic profiles and sleep quality in healthy persons. There could be many reasons, especially that the small sample size of 12 may be insufficient to replicate the results of the other studies. This

could be due to the fact that despite age, gender, and geographical differences, the study cohort may be a homogeneous group of early-adopter highly-interested persons who are already experiencing high sleep quality and better than average favorable genomic profiles for sleep. There may not be a correlation between this particular phenotypic instrument (Pittsburgh Sleep Quality Index) and the SNPs implicated for sleep quality on a general basis as explored in this study. Finally, a great deal of research has focused on genomic linkage to sleep pathologies, and sleep quality in healthy individuals is a new field where additional efforts may lead to replication. However, it is important to notice the potential value of crowdsourced citizen science studies as a resource for attempting to replicate the results of academic studies in additional cohorts, with greater permutation, and also as a potential mechanism for new discovery.

Discussion and Future Perspectives

What is “Awareness”?

MyFinder has the theme of “information environment that gives awareness to a person”. In order to realize the theme, it is necessary to understand what the “awareness” actually is. In this paper, awareness is defined as “to let an individual be conscious of her/his unconscious behavior”. The QS experience and the personal genome experience of the author were described in the third section as awareness was obtained from the viewpoint of the “individual”. Next, the “scientific discoveries by community computing” were described in fourth section as awareness obtained from the viewpoint of the “collective intelligence”.

The awareness can be systematized by the two axes of consciousness/unconsciousness and myself/others. In view of the “Johari window” (a graph model of awareness in interpersonal relationship), which is often referred to in communication psychology and the like, the self can be divided into (1) Open Self (self recognized by the person herself/himself and by others), (2) Hidden Self (recognized by the person herself/himself but not by others), (3) Blind Self (not recognized by the person herself/himself but is recognized by others), and (4) Unknown Self (not recognized by the person herself/himself and others), and the efforts for expanding the region of Open Self and Hidden Self can be considered as the process of bringing about awareness.

MyFinder supports the expansion of the range of self recognized by the person and the self recognized by others by letting an individual know the differences between her/his behaviors, senses, and thinking characteristics (i.e., supports understanding Open Self). Also, in cases where MyFinder is used to become aware of the differences in behaviors, senses, or thinking characteristics, MyFinder gives opportunities for developing mutual understanding by becoming aware of prejudice and sense of

discrimination which have unconsciously been generated. An important aim of MyFinder is to support the process of gaining awareness that allows a deeper mutual understanding.

Social Significance of MyFinder: Design of Future Society

In the Introduction, we raised the question “what if our decisions are in fact controlled by characteristics caused by personal genes and unconscious signals from the society surrounding us to considerable degrees?” In essence, we suggested the following two points:

- We do not know our own minds as well as we think we do. Supporting technology that enables a person to be conscious of her/his unconscious behaviors will become important.
- From now on, research on brain science and genes will make progress to scientifically clarify the differences among individuals. Society should be directed into respecting diversity (the concept of “gifted”: “we are different and that is good”, not toward prejudice and discrimination).

In relation to the question above, the following challenging questions would be raised.

- If decision-making is indeed influenced by genes and unconscious social signals, are “freedom” and “responsibility” merely based on nothing but the illusion of freedom?
- Is it really beneficial for a person to know her/his abilities and characteristics?
- How can a person be conscious of his/her self if he/she is different in comparison to others?
- How can the act of knowing herself/himself be converted into her/his happiness?
- What is essential in order to direct our society towards respecting personal differences as personality, rather than towards discrimination?
- How should we share the personal data in view of the above-identified questions?

These challenges will promote artificial intelligence research and contribute to improvements in the quality of life of human kind.

Acknowledgements

This research was supported by a PRESTO grant for “Proposals and Evaluations of Personal Genome Information Environment using Gene Analysis and Artificial Intelligence Technology” from the Japan Science

and Technology Agency (JST) and is conducted with the tremendous support by Riken Genesis Co., Ltd., Stanford University, DIYgenomics, and many research collaborators.

Contributions

The paper was primarily conceived and written by Kido. Swan contributed to the citizen science study section.

References

- [Ashley 10] Ashley, E.A., Butte, AJ et al. "Clinical assessment incorporating a personal genome." *Lancet*, 375, (9725): pp. 1525-1535 (2010)
- [Bodenmann 12] Bodenmann, S., et al. "Polymorphisms of ADORA2A modulate psychomotor vigilance and the effects of caffeine on neurobehavioural performance and sleep EEG after sleep deprivation." *Br J Pharmacol*, 165(6): pp. 1904-13 (2012).
- [Butte 08] Butte, A.J. "The ultimate model organism." *Science*, 320 (4), pp. 325-327 (2008)
- [Ciarleglio 8] Ciarleglio, C.M., et al. "Genetic differences in human circadian clock genes among worldwide populations." *J Biol Rhythms*, 23(4): pp. 330-40 (2008).
- [Ishida 12] Ishida, T. (Ed.), *Field Informatics*, Kyoto University Field Informatics Research, Springer (2012)
- [Kahneman 11] Kahneman, D., *Thinking, Fast and Slow*. New York, NY: Farrar, Straus, and Giroux. (2011)
- [Kido 11] Kido, T., Genetics and artificial intelligence for personal genome service -MyFinder: Intimate community computing for genetic discovery-, AAAI 2011 Spring Symposium (2011)
- [Kido 12] Kido, T., Self-Tracking Mindfulness Incorporating a Personal Genome, AAAI 2012 Spring Symposium (2012)
- [Kido 13a] Kido, T., Swan, M., Exploring the Mind with the Aid of Personal Genome -Citizen Science Genetics to Promote Positive Well-Being, AAAI 2013 Spring symposium (2013)
- [Landolt 11] Landolt, H.P. "Genetic determination of sleep EEG profiles in healthy humans." *Prog Brain Res*, 193: pp. 51-61 (2011).
- [Minsky 06] Minsky, M. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence and the Future of the Human Mind*, Simon & Schuster, (2006)
- [QS 11] Quantified Self Conference 2011. May 28-29, (2011) <http://quantifiedself.com/conference/Mountain-View-2011/>, Last accessed: Sep 2013.
- [Kido 10] Takashi Kido, *What My Life Will Be Like?: Search for Myself Clarified by Genome*, Hoshinowa-kai (2011)
- [Nova 12] Nova, P., et al. "Modeling caffeine concentrations with the Stanford Caffeine Questionnaire: preliminary evidence for an interaction of chronotype with the effects of caffeine on sleep." *Sleep Med*, 13(4): pp. 362-7 (2012).
- [PG 13] Joel T. Dudley, Konrad J. Karczewski, *Exploring Personal Genomics*, Oxford University Press (2013)
- [Rui 12] Rui Chen, et. al., Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes, *Cell*, volume 148, issue 6, 1293-1307, 16, March, 2012.
- [Venter 10] J. Craig Venter, et. al. An agenda for personalized medicine, *Nature*, vol. 461, 8, October (2009)
- [Kido 13b] Kido, T, Kawashima, M, Nishino, S, Swan M, Kamatani N, Butte AJ. Systematic Evaluation of Personal Genome Services for Japanese Individuals. *Nature: Journal of Human Genetics* (2013)
- [Sasaki 05] Hiroyuki Sasaki, *Introduction to Epigenetics: How the Markings of Calico Cat Are Decided?*, Iwanami Shoten, Publishers (2005)
- [Pentland 13] *Honest Signals: How They Shape Our World*, Misuzu Shobo (2013).
- [Kamatani 10] Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N., Genome-wide association study of hematological and biochemical traits in a Japanese population., *Nat Genet. Mar;42 (3):210-5* (2010).
- [Sandel 10] Michael Sandel, *The Case against Perfection: Ethics in the Age of Genetic Engineering*, Nakanishiya Shuppan (2010)
- [Helen 96] Helen Palmer, *New Enneagrams*, edited by SCAN Communications, (1996)
- [Hayashi 06] Hayashi, T et al. "Laughter regulates gene expression in patients with type 2 diabetes." *Psychotherapy and Psychosomatics*; 75: 62-65. (2006)
- [Bodenmann 12] Bodenmann S, Hohoff C, Freitag C, Deckert J, Re'tey JV, Bachmann V, Landolt HP. Polymorphisms of ADORA2A modulate psychomotor vigilance and the effects of caffeine on neurobehavioural performance and sleep EEG after sleep deprivation. *Br J Pharmacol*. 2012 Mar;165(6):1904-13.
- [Ciarleglio 08] Ciarleglio CM, Ryckman KK, Servick SV, Hida A, Robbins S, Wells N, Hicks J, Larson SA, Wiedermann JP, Carver K, Hamilton N, Kidd KK, Kidd JR, Smith JR, Friedlaender J, McMahon DG, Williams SM, Summar ML, Johnson CH. Genetic differences in human circadian clock genes among worldwide populations. *J Biol Rhythms*. 2008 Aug;23(4):330-40.
- [Landolt 11] Landolt HP. Genetic determination of sleep EEG profiles in healthy humans. *Prog Brain Res*. 2011;193:51-61.
- [Nova 12] Nova P, Hernandez B, Ptolemy AS, Zeitzer JM. Modeling caffeine concentrations with the Stanford Caffeine Questionnaire: preliminary evidence for an interaction of chronotype with the effects of caffeine on sleep. *Sleep Med*. 2012 Apr;13(4):362-7.