Genetics and Artificial Intelligence for Personal Genome Service
MyFinder: Intimate Community Computing for Scientific Discovery

Takashi Kido

RikenGenesis Co., Ltd. and Japan Science and Technology Agency
Taito-ku, Taito, 1-5-1, Tokyo, Japan and 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan
takashi.kido@rikengenesis.jp

Abstract

It is now time to begin the study of personal genome services based on the interdisciplinary theories and technologies of genomics and artificial intelligence (AI). Although recently much attention has been given to personal genome services for realizing personal medicine, little systematic research has been done on their communication and computational aspects for intelligent wellness service in AI communities. We believe that the intelligent personal genome services of the future need to include an understanding of how the knowledge of genetic risk influences people's behavior. This paper proposes the concept of MyFinder, a new framework for realizing an intimate personal genome service with AI technologies. This paper also describes the grand challenge problems of personal genome services that the AI and genomics communities should tackle jointly.

Introduction

We have entered the era of Personal genome. The technological breakthrough and recent drastic cost down of whole human genome sequencing will soon give people an unprecedented opportunity to find their own biological and even psychological compositions. Consumer genetics companies, such as 23andMe, provide genetic testing for over 100 traits and diseases as well as DNA ancestry.

While high-throughput technologies of DNA analyses make it possible to produce amount of personal genome data drastically faster and cheaper, few systematic studies have been done on ideal information environments and people's attitudes about personal genome services.

The purpose of this paper is (1) to propose new research framework for scientific discovery with personal genome communities and (2) to provide grand challenge problems that genetics and AI community should jointly tackle.

In the following sections, we explain the need for studies on personal genome service and propose the concept of MyFinder which is the new research framework for scientific discovery with community computing in personal genome environment. We also describe three grand challenge problems to realize the concept of MyFinder.

The Concept of MyFinder

We believe that the study of personal genome service meets the interdisciplinary theories and practices between AI and genetics. MyFinder is the new research framework in the personal genome environment, which combines the idea of intelligent agent in AI community and the genome wide association studies in genetics community. I show this concept in Fig 1.

One of the goals of MyFinder is to provide personal genome environment which enables us to find our inborn natures, strengths, and talents and to maximize our potential. While traditional personal genome service mainly aim to realize personal medicine such as disease treatment or drug response, MyFinder focuses much more attentions on wellness, psychiatry and behavior sciences.

MyFinder supports the research assumptions on recent genetic switch theory in which our physical, chemical and psychological stress may strongly affects our gene activities. For example, recent some studies report that our laughing activities is very effective for diabetes treatment (Laugher is the best medicine.), which suggests the possibility that the positive mental stress caused by laughing activities may turn on our genetic switch of the cellular activities.

By observing and analyzing our daily behaviors such as diet, sleep, working style, time management, social interactions and individual's preferences, we may monitor our daily physical, chemical and mental stress. Intelligent software agent technologies will be useful for learning the user's behavioral trend and analyzing the stress state.

Another goal of MyFinder is to provide research platform for scientific discovery with community computing. Similar concept has already been implemented by
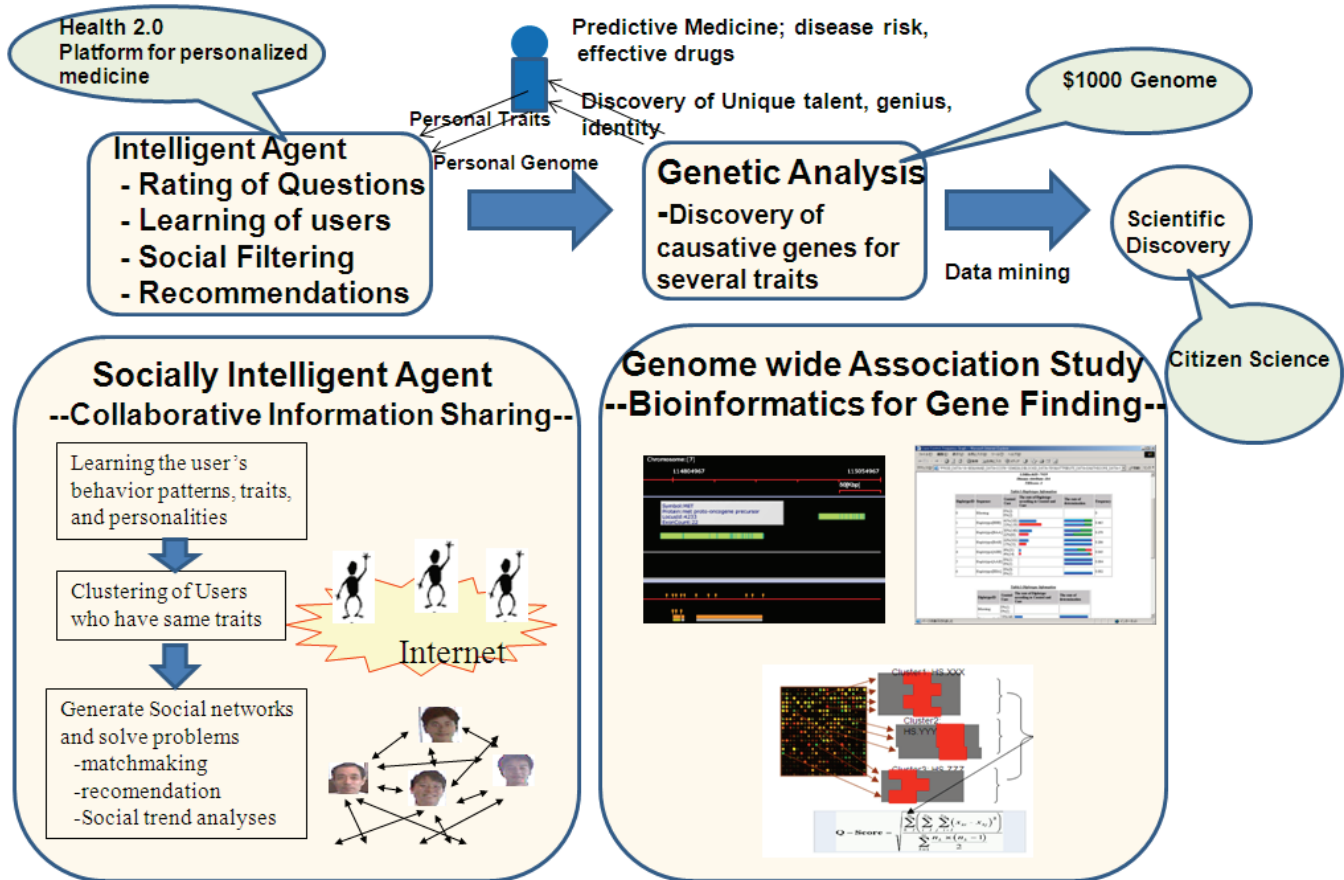
MyFinder

Fig 1. The Concept of MyFinder

23andMe. 23andWe, which is one of the services provided by 23andMe, gives consumers the opportunity to leverage their data by contributing it to studies of genetics. With enough data, 23andWe might produce revolutionary findings that will benefit us all.

While 23andWe requests users to answer some irrevocable surveys for getting user's phenotype data, MyFinder aims to provide more interactive and intimate interface to obtain reliable phenotype data using software agent technology. For example, MyFinder may learn user's personality by monitoring user's daily behavior and suggest some findings to users with psychological or behavior science evidence (for example, based on enneagram theory). This function might be useful for users to rediscover his/her inborn potential.

Grand Challenge Problems for AI in Personal Genome Services

We propose the following three research projects in per-

sonal genome services as challenge problems which AI community should tackle.

(1) Acquisition of reliable personal genome data and reliable genetic risk prediction
Our first challenge is to create the world's trusted source of personal genetic information. Acquiring reliable personal data (genotyping data and sequencing data), reliable phenotype data and reliable knowledge of personal genome research is the key to success. For this purpose, the quality assessment measurement of the personal genome data and innovative algorithms for genetic risk prediction should be developed.

 - Quality assessment, phenotype annotation of whole genome sequencing
The pioneer research on clinical assessment incorporating a personal genome has recently done by Stanford research group with Stanford Steve Quake genome. (Reference 1,3). BGI group in China has been actively doing whole human genome sequencing for Asian people with next generation sequencing technologies. (Reference 2). The challenges for AI communities are reliably and effectively integrating

the huge personal genomes and phenotype annotation data to create trusted database.

-Innovative algorithms for genetic risk prediction

Although Different genetic ethnic groups have different genetic risk (Reference 7), little studies have been done on the risk prediction model for several diseases incorporating several ethic group differences. Machine leaning technologies can be applied for public personal genome data set to learn these differences. Another challenge for AI communities is to develop the sophisticated prediction algorithms for incorporating multiple genes and multiple environment factors. Although current disease association models are based on the genetic statistics theory for mono gene (SNP) and phenotype, the cause and effect models on more complex interactions among genes and environment should be investigated. Probabilistic graphical models based on Bayes theory are promising for this purpose.

(2) Quality analyses of personal genome service and behavior analyses on people's attitude

Even though reliable genetic sources are given, how to report genetic risk information to the people is very important problem. We have to understand how genetic risk knowledge may affect people's behavior. AI community needs to investigate communication aspects on personal genome services. We propose the following two challenges.

-Comparative study on quality of Service for personal genome service

The researchers of personal genome services can learn much from the research history of web search engines. Current situations of personal genome industry are very similar to the 1998's web search industries. Laurence's comparative studies on quality of services of web search engines (such as comparisons of Altavista, Google, Yahoo etc.) published nature and science contribute to find the essential factors of search engines service in quantitative way. Similar research approach can be applied in personal genome industry. (Reference 5, 6)

Recently, Craig Ventor reports the comparative studies on the genetic risk reports of direct consumer firms. (Reference 4). According to this report, for seven diseases, only 50% or less of the predictions of two companies agreed across five individuals. Ventor suggests that genetics community should study markers in all ethnicities and look at people's behavior after tests.

AI community should investigate to find the quantitative factors which will be essential factors for future innovation of personal genome service.

-Behavior pattern analyses on people's attitude for personal genome

Although genomic information becomes increasingly integrated into people's healthcare, little systematic researches has been done on communication aspects in personal genome service. We need to know which types of genomics information are helpful, harmful, and natural. We also need to investigate how the genetic risk knowledge affects people.

We show our idea on people's behavior pattern analyses to investigate the effect of genetic risk knowledge. (Fig. 2). In this example, different type of genetic risk knowledge is shown; one is the risk for Perkinson's disease, the other is the learning ability from errors. The challenges for AI researchers are learning the people's attitude for several genetic risks based on the results of social experiments, classifying the genetic information with several measures (such as credibility, harmfulness, seriousness, usefulness, interest factor) and provide the appropriate information to the appropriate person.

This research is related with the design of socially intelligent agent in AI community. Cultural differences will be also investigated. The findings of these people's behavior pattern analyses can be applied to the design of intimate user interface and new innovative media of human computer interactions.

(3) Data-mining for scientific discovery in personal genome platform

The important goal of MyFinder is to provide research platform for scientific discovery with community computing approach. One of the challenges for AI community is to provide personal genome environment with Web 2.0 and cloud computing technologies.

Recently new models of community based research platform emerges; citizen science, DIYbio, and consumer genomics services. For example, DIYgenomics starts the crowd-sourced clinical trials which replicates, extends, and develops new studies linking genotype with phnotype/behavior/environment in peer cohorts. These studies include disease studies, drug and environmental response studies, and behavioral studies. Designing intelligent and intimate personal genome platform is one of the key to success for these citizen science activities.

Another important challenge for AI communities is to provide data-mining tools for scientific discovery in personal genome platform. Machine learning technologies and effective combinatorial optimization algorithms for large scale data can be applied to find meaningful knowledge from the combination of genomic data, phnotypic biomarker data, self-reported observational data and environmental data. These tools will play important roles for future scientific discovery in community based personal genome environment.

Summary

In this paper, we emphasis the importance of the studies on personal genome service for AI and genetics communities should jointly tackle. We have proposed the concept of
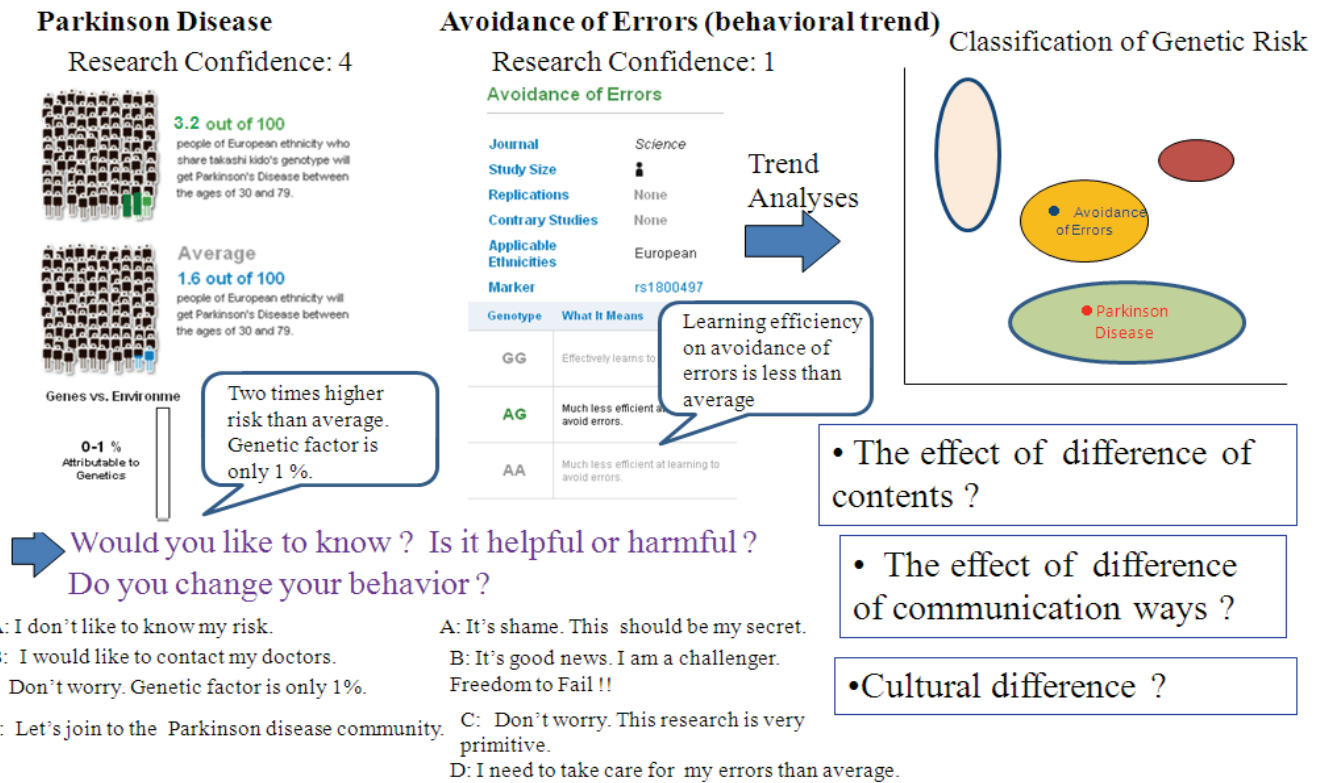
Fig 2. Behavior pattern analyses on genetic risk knowledge

MyFinder; the intimate community computing platform for scientific discovery. We have also provided three grand challenge problems for realizing intimate personal genome service with AI technologies. We hope further discussion will be done in the following workshops.

References

(1) Ashley EA, et.al. , *Clinical assessment incorporating a personal genome.*, Lancet. 2010 May 1;375(9725):1525-35

(2) Jun Wang et. al., *The diploid genome sequence of an Asian individual*, Nature, 2008 Nov 6;456(7218):60-5

(3) Pushkarev D, Neff NF, Quake SR, *Single-molecule sequencing of an individual human genome.*, Nat Biotechnol. 2009 Sep;27(9):847-52. Epub 2009 Aug 1

(4) J. Craig Venter, et. al. *An agenda for personalized medicine*, Nature, vol. 461, 8, October, 2009

(5) Steve Lawrence and C. Lee Giles, *Searching the World Wide Web*, Science 280 (1998): 98-100

(6) Lawrence S, Giles CL., *Accessibility of information on the web*, Nature. 1999 Jul 8; 400 (6740):107-9.

(7) 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. 2010 Mar;42 (3):210-5. Epub 2010 Feb 7