Increasing Social Awareness of Personal Genomics and Bioinformatics

Lucia Bianchi and Pietro Lio'
Studio Bianchi (Firenze, Italy),
The Computer Laboratory,
University of Cambridge (United Kingdom)
luciagbianchi@gmail.com, pl219@cam.ac.uk

Abstract

Being on the Internet poses constants threats to security and privacy. While we are connected and we share information, websites and internet services collect various types of personal data with or without the user consent. It is likely that genomics will merge with the internet culture of connectivity. The growing importance of genomic traffic, the inclusion of sequence data information in electronic health records and the new powerful genome editing technologies urge more efforts in education and social awareness of how biomedical data is stored, processed and transferred through the net.
To achieve more widespread awareness of personal genome and bioinformatics data, we discuss the following targets: 1) the integration of genome data into electronic health records will boost interoperability aspects; this will require more knowledge of bio-ontologies and the data formats used in data sharing. 2) Privacy issues and aspects of user vulnerability will link together bioinformatics aspects of genomics and internet services. In this respect, Internet and Genomics data controllers should be understood as part as a single normative community. 3) The growing success of gamification in health care suggests the importance of adopting gamification in bioinformatics and genome data analysis. We revise the state of art game development used in biomedical fields and Bioinformatics. 4) Finally we discuss the potential role of collective awareness platforms as ad-hoc social networks to promote gamification, knowledge of Bioinformatics used in genome analysis and interpretation (particularly for social-related characteristics), discuss genome privacy related policies and monitor genomic data and net neutrality.

1. Introduction

A 2013 Blog article from Francis Collins stated: "DNA has entered the digital age" [1]. In the near future, genomic data traffic will grow more than other Big Data science such Astronomy, YouTube, and Twitter [2]. This trend will likely expand the role of bioinformatics in human online activities. Development and improvements in Genome technology are moving rapidly [2]. NGS sequencing will probably become as common and cheap as to book a flight. The concern for safety and privacy in genome information sharing has been addressed by scientists leading international genomic projects [3,4]. Less advanced is the

awareness in public administrators, politicians and even less in citizens and therefore the legislation is far behind.
A series of recent papers have stressed the importance and success of teaching bioinformatics in schools [5-9]. The recent Spanish project described at www.sacalalengua.org is an interesting example of teaching genomics and bioinformatics to children. The growing role of Internet in genomic data sharing and new powerful DNA technologies urge a more widespread education on how biomedical data are treated. Although bioinformatics is a technical and specialized field of computer science, it is important that citizens become aware of what ICT and Bioinformatics can do with biomedical data and how to monitor genome data traffic and data controllers.
We believe that there is the need to spread awareness on how genomic and medical information could be transferred and integrated across different formats, procedures and biomedical devices. The integration process is termed interoperability and improvements in interoperability could lead to save money, avoid errors in procedures and recognize iatrogenic effects.
Then we discuss the social aspects of disclosure and vulnerability due to the combined accountability for genomic and other personal online data (for example messages, photos). Games are effective educational supports in many fields; they are still not so much developed in Bioinformatics education. We review state of art gamification in biomedical fields and discuss the potentialities for genome information awareness. The collective awareness platforms are among the most powerful social network frameworks to foster discussion and share knowledge and games on genome data policies and traffic. We list the existing collective awareness platforms and related tools and discuss how those platforms developed by social networks in various fields [10] could inspire similar developments for genomic and bioinformatics data monitoring. We conclude with a vision of how these four themes could act in synergy.

2. Bioinformatics-EHR interoperability

There is a growing attention at the inclusion of genetic information (together with details about the bioinformatics tools used to analyse it) in electronic data records, EHR, [11-16]. This also includes epigenetic data [17]. The effective integration of genomic data into the health care systems will require EHRs to be fully interoperable. Most common formats used in genomic information are FASTA, FASTQ, SAM/BAM, GFF/GTF, BED, and VCF. Researchers in Bioinformatics and Systems Biology have made efforts to standardise ontologies (for example gene and disease ontologies) and data exchange formats in various fields. The biomedical information describes a time and space variant process has multi scale that has multi omic content; it is organised through controlled vocabularies and formats for the exchange of structured data. Bio-ontologies are consensus-based, controlled vocabularies for biological and medical terms and interaction between humans and computers. Addressing the medical comorbidity conditions in the EHR will require to ontologies that take into account the profound wiring of biological

processes in the body. It is noteworthy that ontologies have been used for education on bioontologies [25].
A single human genome sequence constitutes a several-gigabyte data file; a seamlessly interoperability may require standards in genome compression. The repetitive nature of DNA allows data compression and Burrows-Wheeler-Transform has been used in some alignment tools, and also in bzip2. The compression of sequences from high-throughput projects requires more specialised algorithms [19-22].
There is a growing interest in NoSQL databases as being non-relational and generally easy to scale. Nowadays all non-relational databases, and most of the relational databases, support JSON (JavaScript Object Notation). JSON is a self-describing format similar to XML (Extensible Markup Language) but is more compact and tightly integrated into the JavaScript language. The main advantage is that performing specific analyses is no longer limited to a specific location. Instead, all applications can be accessed via devices connected to the Internet, e.g. laptop, mobile phone, or tablet computer. This enhances the user's productivity by having access to relevant data at any time. In most of the architectures, the Human reference genomes, genome annotation data, and clinical trials data are referred to as master data, whereas patient-specific NGS data and EHR are referred to as transactional data. The analysis of the transactional data is the basis for gathering specific insights, e.g. individual genetic dispositions, and for leveraging treatment decision in course of personalised medicine. One important data format is related to the Hadoop Distributed File System (HDFS) which is designed to store very large data sets reliably, and to stream those data sets at high bandwidth to user applications. In a large cluster, thousands of servers both host directly attached storage and execute user application tasks [22-23].
The new data formats coexist with those established several years ago. The most used formats for the exchange of information is still XML that is a tagged format similar to HTML on which web pages are based. XML based mathematical models are deposited in various repositories such as BioModels Database [26], the Physiome Model Repository [27].
The XML protocol is used in the Internet of Things (IoT) and internet of medical things (IoMT). IoT is an infrastructure that administers a hierarchy of communication networks that sense, process, integrate, and distribute information in a demand-driven and controlled way and therefore use many data exchange protocols. Therefore it will be very important in all biomedical fields. In IoT, each device may communicate with each other device; the device data is then transferred to the servers that send back to devices, to control process or to people. The MQTT (Message Queue Telemetry Transport) protocol is a message queuing protocol and it is used for collecting device data and communicating it to servers. The XML-based XMPP (Extensible Messaging and Presence Protocol) protocol is used to connect devices to users (say a device could be accessed using a phone). The CoAP (Constrained Application Protocol) communicate with low-power sensors and devices via the Internet.
The IoMT vision is a complex device ecosystem that ranges from cloud backend services for hospital and big-data analytics to home, public, industrial, and wearable health sensor devices. Wearable health and IoMT devices will

monitor various health parameters in an individual (for example through model checking and run time monitoring software) and maintain data safety and privacy. Devices based on IoMT will help identifying acute episodes, chronic disease management and personal health training.
The diversity of IoMT device implements health data information formats for data storage, integration and interoperability. This includes biomedical semantics used in bioinformatics in order to achieve overall broader interpretable data significance.
The IoMT needs to be implemented in a physical system that could be a verified biomedical cyber-physical system (CPS). CPS is the integration of physical, computation and communication processes (i.e. hardware, middleware, and cyberware) with three main tasks: sensing, processing and networking. Therefore, CPS covers more aspects and concepts than the IoT. The latest technological discoveries in sensors and wireless communications have allowed the design and the building of physiological sensor nodes. These sensors, together with Cloud Computing, would allow CPS healthcare applications in hospitals and in patients home care. The implementation of wireless sensors allows to observe and monitor patients remotely, (for example using model checking and run time monitoring application programs) and take decisions, regardless of the patient's location [28,29]. In typical applications, sensors collect information on patients' health that is then used by runtime monitoring processes. In the case of remote health care monitoring systems, the CPS devices are small embedded applications connected with the human body that capture various body conditions, then process the information locally in real time and send only the relevant output to the cyber space, through internet or via mobile systems. Another aspect that makes healthcare CPS extremely useful is that healthcare conditions could be extremely complex. Rarely patients are affected by a single pathology, in many cases they are affected by comorbidities and multi-morbidities.
Interestingly, several Computer Science and Engineering departments and schools are grouping biomedical CPS and Bioinformatics in the same division; Bioinformatics workshops are often including CPS in the call of papers topics. We should therefore expect more integrate developments. Figure 1 describes the links between of bio-ontologies and XML specifications in (clinical) bioinformatics, IoMT and medical cyber-physical systems. It is likely that the connections will become more dense along with the growth of the number and variety of applications, users, and across the evolution of Internet (Web 2,3,4 in the figure). It is likely that there is a coevolution of formats in biomedical fields and the Web. The Web will pass from a semantic to a cognitive characterisation. Examples of various currently used semantic annotations software (label "Sp") and XML-based standards for exchanging richly annotated data (label "Go"), and CPS/IoMT (label "Pr") are listed in table 1. It is noteworthy that XML-based formats are also enriched with freely available viewers, converters and various tools and libraries. Clearly the bioinformatics interoperability will be greatly extended by the XML protocols across a variety of applications. The citizen should be aware of the circulation of the data provided by the interoperability. The transition from semantic to cognitive may result in higher relevant and personalised information by integrating a variety types of

information, such as TV programs, music. It will implement some level of memory, judgment, and reasoning, to learn the user information preferences adding the ability to act and communicate autonomously, helping in particular in cases of collaborative decision-making and conflict resolution. Requests could be the same or simpler (“send me everyday a report on this topic”) than in the past but the results more relevant and insightful. In figure 2 we show the variety of biomedical information that could be linked together, determining a large scale circulation of different information on a single person and his/her neighbours. The implementation of Web 4- related machine learning and bioinformatics techniques on genomic and social network data will add an impressive level of inference, probably resulting in incidental findings, disclosures and vulnerability. Figure 2 shows how a large number of tables could be linked if they share fields/attributes. The links could be form a probabilistic network that could be easily stored in available memories. For this reason, the genome information transfer should involve patients themselves in multiple layers of consent decision-making process.

Position for Figure 1

Figure 1 illustrates the interdependency of bio –ontologies (GO), specifications (Sp), (clinical) bioinformatics, Internet of medical things and medical cyber-physics with respect to the growth in personal genomes (y-axis), education and social awareness (x-axis). The figure shows also a gradient of technological improvement in the web technology (Web 1.0, Web 2.0: social two-way conversations; Web 3.0: ubiquitous mobility, portability and connectivity, everything to everywhere, philosophy of openness, semantic and intelligent web; Web 4.0: new wave of applications and user interfaces to take advantage of the semantic and ubiquitous web, perhaps further evolving into a cognitive web). The evolution of the web is inspired from the https://mcgratha.wordpress.com/2010/12/30/evolutionary-road/

Position for Figure 2

The figure illustrates the various types of information (categorical attributes) that could be shared online. In green the phenotype, in red the genotype information. The information could be present in an aggregated way in frequency tables that display the count of respondents at the crossing of the categorical attributes. As an example, let’s consider typical tables reporting the number of patients per disease and municipality or disease per age. Several tables could be linked if they share fields/attributes, e.g. “Disease” × “Town” and “Disease” × “Gender”. The fact that marginal row and column totals are preserved makes possible the disclosure attacks of tables reporting aggregate information. Information in linked tables with few respondents is easier to attack.

Here Table 1

4. Social awareness of genome data sharing

Many consortia that include genome centers and hospitals send the message that the medical treatment at personal level could benefit from the genomic data of millions of other people (see for instance www.technologyreview.com/featuredstory/535016/internet-of-dna/)[30-32].
Recent work has focused on incidental findings i.e. findings that are not within the scope of the original objectives of the research,
although these findings are occasionally valuable for patient care [33]; incidental findings highlight the importance of pre-test patient education and having a policy to disclose test results to family members [34].

Personal genome sequencing is becoming cheap as healthy individuals are attracted by offers of whole exome/genome sequence. Other type of omic information such as epigenetic has different facets of privacy. The genome sequence may affect relatives; The epigenetic information may be subjected to incidental findings than the genome information itself. For example the epigenetic data is different for the different tissues and changes with age, disease, inflammation and stress conditions.

Clearly, a new technology is likely to be used also in unforeseen ways (see for instance openplant.org/gosh/), so that private sharing of genomic and other personal information could become frequent. It is important to take a combined view of the accountability of Genomic and Internet data. In other words, it is important to discuss how privacy and security problems of the genomic data will add to unsolved problems about security and privacy of Internet science.
Since its beginning, Internet has created a culture of participation and connectivity that reflects the Internet's potential to create connections and communities and advance democracy. The opportunities given by connectivity drove users to the web but as their life became permeated with social media platforms, they start sharing creative content and move their social and professional activities to an online environment. Therefore, Internet needs to remain open and neutral [35]. Internet neutrality means that everyone sending data should be treated the same by service providers and not charged differently for faster transmission speeds. Although critics of net neutrality focused on the slow down performance at peak times because of too high video or online gaming traffic, net neutrality bias could affect various health services, such as Telehealth and electronic health records exchange. In a culture of connectivity where our social lives are increasingly mediated by mass self-communication, people do not always own the necessary capabilities to optimally interpret and act to acquire an equal position in society [36,37]. This problem is going to stay in the future. In Internet the Culture of connectivity is increasingly mediated by algorithms used in internet media, social media, mobile media, wearable media and then ubiquitous media (for example IoT). Internet science and connective media present large evidences of user disempowerment and an increase of privacy vulnerability. The User is

disempowered if he/she lacks/has lost awareness and control of his/her situation and environment [38]. The user disempowerment depends on the lack of technical knowledge of how mechanisms operate (for example the above mentioned algorithms), as well as on the skills to change them [39].

Data ownership concentration is an additional concern [40]. Online social networks are collecting huge amount of data in the form of text, images and links. Nowadays, the infrastructures of the social networks are often managed by a single company, which could perform user profiles data mining for targeted advertisements, in order to be paid for the investments in storing information.
Most of the online personal information (messages, photos, etc.) may be considered the equivalent of the “phenotype” which could be used to match the genotype information. Young individuals update very frequently the online information; this information could be correlated with the bioinformatics analysis of the complex genomic traits. The expansion of genetic data sharing and the connectivity culture will generate user vulnerability and the need for a social dimension of bioinformatics.
There are online tools to retrieve personal information from various online source, examples are Abika.com, USSearch.com, 123people.com, PeopleSearch.net, PeopleFinder.com, AnyWho.com and social network aggregator web sites such as Lifehacker.com, Spokeo. com, Spoke.com and Intelius.com. The extension of these tools to biomedical information could mean the aggregation of social network information (such as those from Facebook) and those from Biomedical databases.
This increases the likelihood of re-identification because a disease that is clearly visible, and provides information about the likely geographical location, or ethnicity. Anonymised genome data can also sometimes be re-identified by other ways, such as surname inference of well annotated databases. De-identification procedure often consists of removing identifiers such as images, names, date of birth, email, phone numbers, etc. The cross-data mining of social networks and rare or infectious disease information such as malaria, enteritis may provide some hints on the exposome of an individual [41] or the individual’s likely ancestry or geographical location [42-44]. Severe conditions identified through images or text information are unlikely to be kept private once symptomatic.
Therefore, Internet and Genomics data controllers should be understood as part as a single normative community. The normative dimension is largely missing and models of strict liability are not frequent in our society. Despite lack of prioritisation in the political agenda, there is a need to make a serious effort into accountability and indeterminacy for the genomic data in Internet and to understand how the accountability depends on net neutrality policies in the galaxy of internet providers and the few centralized data centers. The liability of data controllers and how this liability scales according to technological possibilities is becoming a key issue in order to safeguard the balance between reducing the user control over his/her online activities while at the same time unburdening the users of part of their responsibility.
Vulnerability is often distinguished in an external, structural aspect, related to “exposure” and internal aspect, related to perception and coping capabilities

and action of the individual to overcome or at least mitigate negative effects [45]. Vulnerability has a close link with user disempowerment because online user practices have social media behaviors that are more persistent in time, with a very extensive geographical reach and are often picked up by unwanted receivers anywhere in the world.
The internal side of vulnerability clearly depends on the awareness, technical skills and social, cultural and psychological context of the users. The external side of vulnerability depends on the context in which a technology comes into existence. Technology is often not neutral, but introduced as or transformed to be biased in one way or another. However people can have different unforeseen readings and usage of these technologies [46].

An important aspect of the vulnerability in managing online privacy, is the combination of exposure by media technologies and the coping capabilities of users. The coping capacity is different between young and mature people, as young are usually quicker in incorporating the digital media into their everyday life.
The norms and values of what is morally right or wrong are not fixed but are exactly the elements that are at stake [47]. Platform owners, shifting cultural perspectives, governance decisions, policy priorities and commercial reasons all co-determine the online space evolution in which young people define their moral perspective on what they should or should not disclose.
The characteristics and traffic of genomic data may amplify the two exposures. The genome is a unique type of information and the sense of exposure could be long lasting and may require psychological assistance. Moreover, the patient rights or hopes could contrast the privacy of healthy siblings. Given the impressive growth of genomic traffic, the policy of Internet data stewardship and control and internet neutrality should be central items in the agenda of both internet and genomics policy makers. Increasing Bioinformatics awareness will help framing the process.

4. State of art opportunities for gamification of bioinformatics

Games are effective learning environments because the player has lot of control, is motivated for out-of-the-box thinking and there is no punishment for failure. Examples of game development software are Pygame, Unity, Gideros, Minecraft, Scratch (which runs on Raspberry pi architectures; www.raspberrypi.org/learning/getting-started-with-minecraft-pi/) [48-51].

There is a growing interest in using games in education [52], health [53,54]. Games are successfully used for bioinformatics purposes; for instance foldit is an online puzzle video game about protein folding (http://fold.it/portal/; games.cs.washington.edu/DNA_Game_for_Rosetta/DNA.html).
Many games are inspired by 3D Tetris where the user tries to fit 3D blocks together, filling the empty spaces or by Minecraft where the user tries to build your own protein starting from a scaffold.

The users are motivated by the sense of purpose of contributing to science, even without background in bioinformatics, genetics or biochemistry. The encouragement of competition is achieved through different rewards, motivations and social interactions: the user tries to fold their version of the protein better than everyone else to get the highest score. Players can form teams to work together. So individual players can start folding the protein, then they can share their solution with other members (social interaction) so the whole group gets credits for what each member has achieved. There are no time-critical kinds of things, other than approaching the deadline for a particular puzzle closing. Instances of gamification could be based on Minecraft-related games. For example bioinformatics data (e.g. multi omics) and clinical data could be visualised as landscapes or buildings in Minecraft scenarios. Adding or changing the various parts of a building (roof, number of rooms, deep fundaments) could represent operations (for example integration, calibration) on different types of data. The social interaction in such games, contains already ingredients of community awareness platforms. We could also extend gamification to EHR. For example, EHR of neurodegenerative diseases may require the patient to use a self reported diary system to annotate precipitants. Diary reporting is often seen as a boring activity and missing data or cheating behaviors are known. Visual note-takers help people to visualise their thinking and can also be part of a personal knowledge base system; an example is Mindjet.
Medical studies have shown that a visual note taker can improve efficiency up to 15% over conventional note-taking [55]. This tool will enable patients and those supporting them to build tree-like and other diagrams to organise and visualise information. It will use a library of icons to express/represent with different sizes, and position on a diagram, events, people, states of minds, feelings, times etc. to be considered as potential precipitants independently or associated with the seizure episodes. This information could then be exported in various formats to electronic health records and could be further edited by a physician to prepare the EHR. A motivating education environment could be built by using social awareness software. Here the gamification could play a central role in providing the environment to be combined with the note taking tool.

Here Table 2

5. Community awareness platforms and tools

The increasing availability of cheap/garage DNA engineering equipment (see for instance), the growing number of large national genome sequencing projects push for more pervasive solutions to achieve a good level of social awareness on genomics. We believe that social awareness platform could represent part of the answer. Education and social awareness are fundamental for a sustainable relationship between society and science and technology. There is a lack of education towards how genomic data could be stored, analysed, transferred, used and manipulated. This is certainly the domain of Bioinformatics but it is subtly connected with the privacy and accountability of internet data. Citizens

need to be aware of the trade-off between the accountability and the anonymity, and how decentralized privacy-enhanced systems, privacy by design, and reputation systems can protect against online criminal activities. The European countries have started funding programs in Collective Awareness Platforms for Sustainability and Social Innovation. The program in Digital Social Innovation will develop ways to cope with sustainability challenges that could have a measurable global impact. In other words, the funding is directed towards building new ecosystems aimed at supporting an open community of participatory innovators and users. The main idea is to produce collective intelligence in important sustainability areas by using knowledge networks, emerging network technologies, open hardware and data. See also https://ec.europa.eu/digital-single-market/news/22-new-caps-projects-horizon-2020

for recent projects. This important effort in building citizen awareness platforms could perhaps develop also in the direction of genomic data and technologies awareness and help empower people in their online life, monitoring net neutrality and genomic data stewardship. Collective awareness platforms could address in an effective way the privacy issues related to genome data and the social exchange of information about traffic of genomic data for research and medical purposes. This type of platform could represent a different type of resource that includes different social, scientific and entrepreneurial actors and communities, i.e. a decentralized, bottom-up social platform driven by local communities. There are several existing collective intelligence and awareness platforms and available software tools (see table 3). Recent initiatives such as “Patients like me” have a profound impact in motivating patients to face proactively their condition and have stimulated positive reactions from physicians and health resource administrators.

The use of collective awareness platforms will result in raising individual awareness on information exposure and developing citizen empowerment with respect to genomic data ownerships. While empowerment is a way to allow individuals to make their own informed decisions about sharing their data, the reality often translates to a shift in responsibility from entities that collect, store and process the data to individuals having to make decisions about the efficacy of concealing or revealing personal information. A community awareness network (CAN) or a Digital Social Platform (DSP) is not simply a community-driven service provider, but a common network infrastructure built, owned, managed and used by a community in common. A CAN is socially inclusive as it is open for participation. CAN members promote new connections and provide all kinds of support to new members to help extend the network, thus enabling widespread participation in other collective digital platforms. CANs are run by communities that have a deep and critical understanding and interests in the problems of the current genome data and related technologies, they see limitations and the perils of loss of citizen control.

Operatively, this type of platform will provide means for social networking application that will implement gamification and other motivational methods to foster user engagement and will enable the actual coordination mechanisms for the social resource exchange. It will provide open ICT and bioinformatics

educational resources to support the adoption of sustainable responses in the domain of network neutrality measurements. See table 3 for existing CAN software platforms that could inspire similar ones in genome data awareness.

Here Table 3

6. Conclusions

In this review article we focus on the social and educational challenges of Bioinformatics in the context of the growth of the internet traffic of the genomic sequences and the evolution of internet itself.
Internet has nurtured a revolution in connectivity and social networks. Genomics is also nurturing a new technological revolution but also a social one. It is likely that privacy and security problems of the genomic data transfer over the Internet will add to unsolved problems about security and privacy of Internet. In other words: "Two wrongs do not make a right". This paper highlights that education in bioinformatics will have a growing importance with the increase genome transfer over Internet. It also points that genomics and Internet should be considered together when discussing the privacy and security issues.
Building a puzzle is often an example of participatory activity among children. Internet and Genomic data policy makers, biomedical communities, public and private sectors could build a puzzle together and they should first develop a shared vision about what is the puzzle.
In order to achieve this target, we have identified different aspects of technical and social awareness that need to be raised. First aspect is the awareness of the genome information mobility and circulation due to the increase interoperability of the formats. Second aspect is the social vulnerability due to the genome and Internet privacy weaknesses. We propose gamification as an important means to spread Bioinformatics and genomic education. Digital social platforms implemented for collective awareness in different areas could be also developed for the education and gamification activities. Therefore, bioinformaticians could play a meaningful role in mitigating the future risk of digital-genomic divide.

7. References

1) Collins, F. Your DNA enters the digital age: Q&A with NIH director. USATODAY 3:20 p.m. EDT May 15, 2014. http://www.usatoday.com/story/opinion/2014/05/14/nih-director-health-digital-dna-cancer-flu-column/9103965/

2) Stephens, ZD, Lee SY, Faghri F, et al. Big Data: Astronomical or Genomical? PLoS Biol 2015, 13(7): e1002195. doi:10.1371/journal.pbio.1002195

3) Carroll, D. Charo, R. The societal opportunities and challenges of genome editing. Genome Biology, 2015,16:242.

4) Calvert, J and Martin, P. The role of social scientists in synthetic biology. Science & Society Series on Convergence Research EMBO Rep. 2009, 10(3): 201–204. doi: 10.1038/embor.2009.15

5) Form, D and Lewitter, F., Ten Simple Rules for Teaching Bioinformatics at the High School Level, PLoS Comput Biol. 2011 Oct; 7(10): e1002243. doi: 10.1371/journal.pcbi.1002243

6) Machluf Y, Yarden A. Integrating bioinformatics into senior high school: design principles and implications. Brief Bioinform. 2013 Sep;14(5):648-60. doi: 10.1093/bib/bbt030. Epub 2013 May 10.

7) Lim SJ, Khan AM, De Silva M, et al. The implementation of e-learning tools to enhance undergraduate bioinformatics teaching and learning: a case study in the National University of Singapore. BMC Bioinformatics. 2009 Dec 3;10 Suppl 15:S12. doi: 10.1186/1471-2105-10-S15-S12.

8) Wightman B, Hark AT. Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. Biochem Mol Biol Educ. 2012 Sep-Oct;40(5):310-9. doi: 10.1002/bmb.20637.

9) Magana AJ, Taleyarkhan M, Alvarado DR, et al. A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research. CBE Life Sci Educ. 2014, 13(4):607-23. doi: 10.1187/cbe.13-10-0193.

10) Sestini, F. Collective Awareness Platforms: Engines for Sustainability and Ethics. In: IEEE Technology and Society Magazine Winter. 2012, 54-62.

11) Joseph L. Kannry & Marc S. Williams Integration of genomics into the electronic health record: mapping terra incognita Genetics in Medicine (2013) 15, 757–760 doi:10.1038/gim.2013.102

12) Bonney W. Is it appropriate, or ethical, to use health data collected for the purpose of direct patient care to develop computerized predictive decision support tools? Stud Health Technol Inform. 2009;143:115-21.

13) Joseph L. Kannry & Marc S. Williams Integration of genomics into the electronic health record: mapping terra incognita Genetics in Medicine (2013) 15, 757–760 doi:10.1038/gim.2013.102.

14) Nishimura AA, Tarczy-Hornoch P, Shirts BH. Pragmatic and ethical challenges of incorporating the genome into the electronic medical

record. Curr Genet Med Rep 2014 Dec 1. 2(4) 201-211.

15) Shirts BH, Salama JS, Aronson SJ, Chung WK, Gray SW, Hindorff LA, Jarvik GP, Plon SE, Stoffel EM, Tarczy-Hornoch PZ, Van Allen EM, Weck KE, Chute CG, Freimuth RR, Grundmeier RW1, Hartzler AL, Li R, Peissig PL, Peterson JF, Rasmussen LV, Starren JB, Williams MS, Overby CL. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. J Am Med Inform Assoc 2015 Jul 3.

16) Wei WQ1, Denny JC Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med 2015 (1) 41.

17) Goebel G, Pfeiffer KP, Schabetsberger T, Kalozy C, Fiegl H, Leitner K. Relevance and management of methylation data in electronic health records. Stud Health Technol Inform. 2009;150:135-9.

18) Crockford D (2006) RFC4627: The application/json Media Type for JavaScript Object Notation (JSON). http://www.ietf.org/rfc/rfc4627.txt. Accessed Sep 23, 2013

19) Saha S, Rajasekaran S. ERGC: an efficient referential genome compression algorithm. Bioinformatics. 2015 Nov 1;31(21):3468-75. doi: 10.1093/bioinformatics/btv399. Epub 2015 Jul 2.

20) Deorowicz S, Danek A, Niemiec M. GDC 2: Compression of large collections of genomes. Sci Rep. 2015 Jun 25;5:11565. doi: 10.1038/srep11565.

21) Wandelt S, Leser U Sequence Factorization with Multiple References PLoS One. 2015 Sep 30;10(9):e0139000. doi: 10.1371/journal.pone.0139000. eCollection 2015.

22) Hasso Plattner (Editor), Matthieu-P. Schapranow (Editor) High-Performance In-Memory Genome Data Analysis: How In-Memory Database Technology Accelerates Personalized Medicine. Springer; 2014 edition (18 Oct. 2013)

23) Wang C, Dai D, Li X, Wang A, Zhou X. Accelerating Computation of Large Biological Datasets using MapReduce Framework. IEEE/ACM Trans Comput Biol Bioinform. 2016 Apr 5.

24) Frédéric Achard, Guy Vaysseix, and Emmanuel Barillot. XML, bioinformatics and data integration. Bioinformatics (2001) 17 (2): 115-125 doi:10.1093/bioinformatics/17.2.115

25) Amith M., Gong Y., Cunningham R., et al. TITOLO Journal of Biomedical

Semantics. 2015, 6:23 DOI 10.1186/s13326-015-0016-2

26) Juty, N, Raza Ali, Glont M, et al., BioModels: Content, Features, Functionality and Use. Pharmacometrics & Systems Pharmacology 2015, 4, 2,55–68

27) Yu T., Lloyd C.M., Nickerson, D.P., et al., The Physiome Model Repository 2. Bioinformatics, 2011 27(5): 743-744 doi:10.1093/bioinformatics/btq723.

28) Wang, J. Abid, H., Lee, et al., A secured health care application architecture for cyber-physical systems. Control Engineering and Applied Informatics , 2011, 13, 101-108.

29)Milenkovic, A, Otto, C and Jovanov E, Wireless sensor networks for personal health monitoring: issues and an implementation. Computer Communications , 2006, 29, 13-14, 2521-2533.

30) Dyke S, Cheung W, Joly Y, et al. Epigenome data release: a participant-centered approach to privacy protection. Genome Biology 2015, 16 :142.

31) McGuire AL, Beskow LM. Informed consent in genomics and genetic research. Annu Rev Genomics Hum Genet. 2010;11:361-81. doi: 10.1146/annurev-genom-082509-141711.

32) Henderson GE, Wolf SM, Kuczynski KJ, Joffe S, Sharp RR, Parsons DW, Knoppers BM, Yu JH, Appelbaum PS. The challenge of informed consent and return of results in translational genomics: empirical analysis and recommendations. J Law Med Ethics. 2014 Fall;42(3):344-55. doi: 10.1111/jlme.12151.

33)Roche MI, Berg JS. Incidental Findings with Genomic Testing: Implications for Genetic Counseling Practice. Curr Genet Med Rep. 2015;3(4):166-176. Epub 2015 Aug 25.

34) Rong Chen et al Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nature Biotechnology (2016) doi:10.1038/nbt.3514

35) Bijlsma RM, Bredenoord AL, Gadellaa-Hooijdonk CG, Lolkema MP, Sleijfer S4, Voest EE, Ausems MG6, Steeghs N. Unsolicited findings of next-generation sequencing for tumor analysis within a Dutch consortium: clinical daily practice reconsidered. Eur J Hum Genet. 2016 Apr 13. doi: 10.1038/ejhg.2016.27.

36) Chambers R. Vulnerability, Coping and Policy IDS Bulletin 2006, 37, 4, 33–40

37) Spain, SL., Pedroso, I, Kadeva N., et al. A genome-wide analysis of putative functional and exonic variation associated with extremely high intelligence. Molecular Psychiatry advance online publication 4 August 2015; DOI: 10.1038/mp.2015.108

38) Sanderson, SC, Linderman M., Suckiel, S.A, et al. Motivations, concerns and preferences of personal genome sequencing research participants: Baseline findings from the HealthSeq project. European Journal of Human Genetics advance online publication 3 June 2015; doi: 10.1038/ejhg.2015.118

39) Powell, A. and Cooper, A. Discourses of Net Neutrality: Comparing Advocacy and Regulatory Arguments in the US and the UK The Information Society 2011,27: 311-325

40) Heyman, R., De Wolf, R. & Pierson, J. Evaluating social media privacy settings for personal and advertising purposes, in Info - The journal of policy, regulation and strategy for telecommunications, information and media, 2014, 16 (4), 18-32.

41) Pierson, J. Interdisciplinary perspective on social media, privacy and empowerment: The role of Media and Communication Studies in technological privacy research, in O'Hara, Kieron David, Scott L., De Roure, David and Nguyen, M.-H. Carolyn (Eds.) Digital Enlightenment Forum Yearbook 2014 – Social networks and social machines, surveillance and empowerment, Amsterdam: IO Press

42) Pierson, J. Online privacy in social media: a conceptual exploration of empowerment and vulnerability, in Communications & Strategies (Digiworld Economic Journal), 2012, 4thQ (88), 99-120.

43) Zimmerman MA, Rappaport J. Citizen participation, perceived control, and psychological empowerment. Am J Community Psychol. 1988 Oct;16(5):725-50.

44) van Dijck, J. The culture of connectivity: a critical history of social media. Oxford: Oxford University Press, 2013, 228

45) Lawrence N. 2015. Beware the rise of the digital oligarchy http://www.theguardian.com/media-network/2015/mar/05/digital-oligarchy-algorithms-personal-data

46) Chambers R. Vulnerability, Coping and Policy IDS Bulletin 2006, 37, 4, 33–40.

47)R. R. Dunn, T. J. Davies, N. C. Harris, M. C. Gavin. Global drivers of human pathogen richness and prevalence. Proceedings of the Royal

Society B: Biological Sciences, 2010; DOI: 10.1098/rspb.2010.0340.

48) Jones DP, Sequencing the exposome: A call to action. Toxicol Rep. 2016; 3:29-45.

49) Bizimana JP1, Kienberger S, Hagenlocher M, Twarabamenye E. Modelling homogeneous regions of social vulnerability to malaria in Rwanda. Geospat Health. 2016 Mar 31;11(1 Suppl):404. doi: 10.4081/gh.2016.404.

50) Bowie C, Campbell M, Beere P, Kingham S. Social and spatial inequalities in Rotaviral enteritis: a case for universally funded vaccination in New Zealand. N Z Med J. 2016 Mar 11;129(1431):59-66

51) https://www.raspberrypi.org/learning/getting-started-with-scratch/

52) https://www.raspberrypi.org/learning/getting-started-with-minecraft-pi/

53) http://www.gamefromscratch.com/post/2015/12/21/Gideros-Now-Runs-on-Raspberry-Pi.aspx

54) http://www.pygame.org/hifi.html

55) http://money.cnn.com/2016/01/19/technology/microsoft-minecraft-education/index.html

56) Boulos KMN, Gammon S, Dixon MC, et al. Digital games for type 1 and type 2 diabetes: underpinning theory with three illustrative examples. JMIR Serious Games. 2015 Mar 18;3(1):e3. doi: 10.2196/games.3930.

57) Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J. 2014 Jul 1. pii: 1460458214537511.

58) Farrand P, Hussain F, Hennessy E. (2002) The efficacy of the 'mind map' study technique. Medical Education 36: 426-431.

Keywords
Personal genomics
Bioethics
XML protocols
Gamification
Collective awareness platforms

Key points of this review

1) Personal genomics has a growing social dimension that requires widespread awareness on vulnerability.
2) Awareness on the interoperability of Bioinformatics bio-ontologies and

XML protocols with Internet of Medical Things, cyber-physical devices and body sensor networks.
3) Privacy and security problems of the genomic data will add to unsolved problems about security and privacy of Internet.
4) Participatory activities such as Gamification and collective awareness platforms may mitigate the risk of digital and genomic divide.

Biographical Note

Bianchi Lucia is an Italian lawyer. She has interest in ethical, privacy and legal aspects of personal genomics. She has published various papers in peer reviewed conference and journals (for example Briefings in Bioinformatics), educational journals (for example in the Italian edition of Le Scienze) and has given talks at summer schools in genomics**.**

Pietro Lio' is a Reader in Computational Biology at the Computer Laboratory, University of Cambridge. His affiliations also include the Cambridge Computational Biology Institute. His work spans Machine Learning and computational models for health Big Data, personalised medicine research, multi-scale/multi-omic/multi-physics modelling and data integration methods**,** with more than 230 peer reviewed papers. He has edited several books and given over 20 keynote talks.

This is important given that XML-based specifications are present in other medical technologies such as Internet of Medical Things, cyber-physical devices and body sensor networks. This interoperability will facilitate data sharing .

bioinformatics roles that couple technical and social awareness aspects of genomics (section 2) and the

Rong Chen et al Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nature Biotechnology (2016) doi:10.1038/nbt.3514

Naveed M, Ayday E, Clayton EW, Fellay J, Gunter CA, Hubaux JP, Malin BA, Wang X. Privacy in the Genomic Era. ACM Comput Surv. 2015 Sep;48(1). pii:6.
A typical architecture consists of application, platform, and data layer. The application layer consists of special purpose applications to answer medical and research questions. Analysis and processing of data are performed within the platform layer eliminating time-consuming data transfer.

builds apps that help users collect and browse their genomic and other healthcare information on mobile devices like iPhones and iPads