

DEBATE

Open Access

# Patient-controlled encrypted genomic data: an approach to advance clinical genomics

Yannis J Trakadis\*

## Abstract

**Background:** The revolution in DNA sequencing technologies over the past decade has made it feasible to sequence an individual's whole genome at a relatively low cost. The potential value of the information generated by genomic technologies for medicine and society is enormous. However, in order for exome sequencing, and eventually whole genome sequencing, to be implemented clinically, a number of major challenges need to be overcome. For instance, obtaining meaningful informed-consent, managing incidental findings and the great volume of data generated (including multiple findings with uncertain clinical significance), re-interpreting the genomic data and providing additional counselling to patients as genetic knowledge evolves are issues that need to be addressed. It appears that medical genetics is shifting from the present "phenotype-first" medical model to a "data-first" model which leads to multiple complexities.

**Discussion:** This manuscript discusses the different challenges associated with integrating genomic technologies into clinical practice and describes a "phenotype-first" approach, namely, "Individualized Mutation-weighted Phenotype Search", and its benefits. The proposed approach allows for a more efficient prioritization of the genes to be tested in a clinical lab based on both the patient's phenotype and his/her entire genomic data. It simplifies "informed-consent" for clinical use of genomic technologies and helps to protect the patient's autonomy and privacy. Overall, this approach could potentially render widespread use of genomic technologies, in the immediate future, practical, ethical and clinically useful.

**Summary:** The "Individualized Mutation-weighted Phenotype Search" approach allows for an incremental integration of genomic technologies into clinical practice. It ensures that we do not over-medicalize genomic data but, rather, continue our current medical model which is based on serving the patient's concerns. Service should not be solely driven by technology but rather by the medical needs and the extent to which a technology can be safely and effectively utilized.

**Keywords:** Genomic, Genome, Exome, Encrypted, Sequencing, Clinic, Integration, I-MPOS, I-MPOSE

## Background

The revolution in DNA sequencing technologies over the past decade has made it feasible to sequence an individual's whole genome at a relatively low cost [1], [2]. The potential value of the information generated by genomic technologies for medicine and society is enormous. However, personal genomic testing may not be ready for widespread clinical use [3]. Analysis, interpretation and management of the data generated from exome sequencing (ES), which targets slightly more than 1% of the entire genome, have already proven to be a challenge in

the research context for mendelian disorders [4]. One can only imagine the level of complexity that will result from trying to incorporate whole genome sequencing (WGS) into *clinical* practice, especially in dealing with common diseases characterized by complex inheritance. Nonetheless, the low cost of WGS may encourage its early adoption creating "a million dollar headache" [5].

First, the extent to which it is possible to obtain meaningful patient informed-consent to large-scale genomic analysis has been called into question [6], [7]. For instance, genomic approaches increase the chance of discovering incidental findings or results with uncertain clinical significance. Mutations with incomplete penetrance, mutations in novel genes with unknown

Correspondence: john.trakadis@mail.mcgill.ca  
Department of Medical Genetics, Montreal Children's Hospital-McGill  
University Health Centre, 2300 Tupper, Montreal, QC, Canada

function, and genetic variants responsible for very small increases in disease risk are typical challenging examples. This information is difficult to conceptualize and can generate emotional distress over disease risk even among healthy individuals [8]. Hence, during counseling, the potential benefits for individuals need to be weighed against the potential harm and the individuals' right not to know.

Currently, genetic risk assessment is restricted to individuals at increased risk based on family history or clinical presentation, which may be indirectly ensuring the necessary genomic background for the pathogenic role of a variant. Recent reports suggest that on average, each person is heterozygous for approximately 50–100 variants classified by the Human Gene Mutation Database (HGMD) as causing inherited disorders and of approximately 250 to 300 loss of function variants in annotated genes [9]. With what degree of certainty can we predict the impact of even known pathogenic variants when they are identified as incidental findings [10], [11]? Moreover, data from personal genome sequencing is a powerful personal identifier challenging the traditional mechanisms of protection. This raises concerns about privacy, confidentiality and potential genetic discrimination [12].

Responsibly managing the very large amounts of genetic information produced is another key challenge to overcome [13]. As stated above, the significance of the data generated from ES/WGS varies from clearly clinically relevant (e.g. "pathogenic mutations") to data of unknown clinical significance ("variants of unknown significance", VUS). The evaluation and management of novel variants identified using ES/WGS require substantial time and cost. Expertise is needed for analysis of the complex genomic data but there are not enough clinical geneticists to interpret results from wide-spread whole-genome sequencing and to provide follow-up information and clinical care [14,15]. Moreover, return of genomic results requires substantial financial resources for genetic counselling of patients, particularly in the context of the dynamic, continuously evolving, nature of the interpretation of ES/WGS results [16]. The practical challenge of the duty to re-contact patients as knowledge changes over time is already experienced in medical genetics secondary to the clinical implementation of array-comparative genomic hybridization (array CGH) and research studies using ES. This will be exponentially more complex in the context of large-scale *clinical* applications of ES, and especially WGS [17].

Overall, premature integration of ES/WGS into clinical care may lead to a cascade effect [18] and strain the health-care resources in a disproportionate, unjustified fashion [19]. The immediate, actionable, applications of the information derived from genomic technologies are

limited at present time [20]. Moreover, approximately 20% of all human genes and, thus, many of the genes with immediate clinical application have patents on them [21]. Reporting mutations identified by genome sequencing in such genes has not been resolved (e.g. the Prometheus and Myriad Genetics cases) [22], [23].

It appears that medical genetics is shifting from the present "phenotype-first" medical model to a "data-first" model which leads to multiple complexities (see Table 1). To accelerate the integration of ES/WGS into clinical practice a substantial shift in thinking is required. This manuscript outlines a novel approach which can help address the above-mentioned challenges. The described "phenotype-first" medical model proposes that ES/WGS could incrementally be implemented in day-to-day clinical practice in the immediate future, as long as our focus remains on diseases which, at the time of the evaluation, have been genetically well characterized.

## Discussion

### Overview of current clinical genetic practice

In current clinical practice, in order to identify the genetic variant responsible for a patient's disease, first there

### Table 1 Challenges of integrating ES/WGS in clinical practice

---

#### Meaningful patient informed-consent may not be feasible

*Complex medical and social implications of the test results,*

*Possibility of incidental findings,*

*Multiple findings of uncertain clinical significance,*

*Insurance companies do not reimburse pre-testing counselling for ES/WGS*

*Multiple issues to discuss leading to prohibitive requirements in time & resources*

#### Potential emotional distress over disease risk even among healthy individuals

##### Genomic information is a powerful personal identifier

*Raising concerns about privacy, confidentiality, genetic discrimination*

##### Very large amounts of genetic information generated

*Limited number of clinical geneticists for data interpretation and clinical care*

*Substantial time and cost for data analysis and genetic counselling*

*Dynamic/evolving nature of the interpretation as new knowledge is gained*

*Duty to re-contact patients as knowledge changes over time*

#### Shift from the present "phenotype-first" medical model to a "data-first" model.

*Currently, genetic risk assessment is restricted to individuals at increased risk based on family history or clinical presentation (ensuring the necessary genomic background), or newborn screening programs meeting specific criteria.*

*Can we predict the impact of even known pathogenic variants outside this setting?*

*Is there adequate evidence to suggest that a departure from these standards will be beneficial to society?*

---

is recognition of a more or less specific phenotype followed by molecular testing of a series of relevant genetic loci, individually or in small sets [24]. There exist a variety of online databases which aim to catalogue all known diseases with a strong genetic component (e.g. mendelian genetic syndromes) and link, at least some of, these diseases to the relevant genes in the human genome. A very well known example of such a database which is freely available is OMIM [Online Mendelian Inheritance in Man; <http://omim.org/>]. Typically, to aid in the diagnosis of rare genetic syndromes, a physician would perform a database search using patient-specific information derived from the medical evaluation performed in clinic. The clinician would then verify the search-results and clinically prioritize genetic testing for the disease that he/she considers most likely. Clinical genetic tests are often very expensive, so typically only one gene, or occasionally a group of pre-selected genes typically responsible for one disease, will be analyzed at a given time. If the first test is negative, sequential testing of the genes responsible for the other, less likely but still clinically suspected, genetic diseases usually follows.

Once a genetic change is identified, the physician evaluates the variant for pathogenicity. Some genetic variants can be unequivocally interpreted as pathogenic or normal variants on the basis of extensive clinical experience [25]. In other cases, clinical experience is insufficient and pathogenicity is inferred based on the variants' predicted effect on protein stability, function, or conservation. Examples of different criteria used in the clinical interpretation of a genetic variant identified by sequencing include: (1) whether the identified variant has been listed as a non-pathogenic variant, (2) whether the variant has been previously reported to cause disease in the literature and/or databases listing such variants, (3) the nature of the mutation: e.g. a silent or a nonsense mutation, (4) the position affected and whether this position is well-conserved across species, and (5) the presence or absence of the mutation in affected or unaffected family members. Computational inferences, although very useful, are not always reliable. Clinical experience shows that even mutations predicted *in-silico* with high-confidence to be deleterious may ultimately have no major consequences [26], [27].

At present, genomic sequencing technologies are mainly limited to research. They are typically used to identify the mutant gene which could explain a specific disease not-yet genetically characterized. Criteria similar to the ones mentioned above for the interpretation of clinical test results are used to evaluate the pathogenicity of the multiple genetic variants identified by ES/WGS and their possible role in the phenotype under investigation. ES has proven to be a powerful research tool

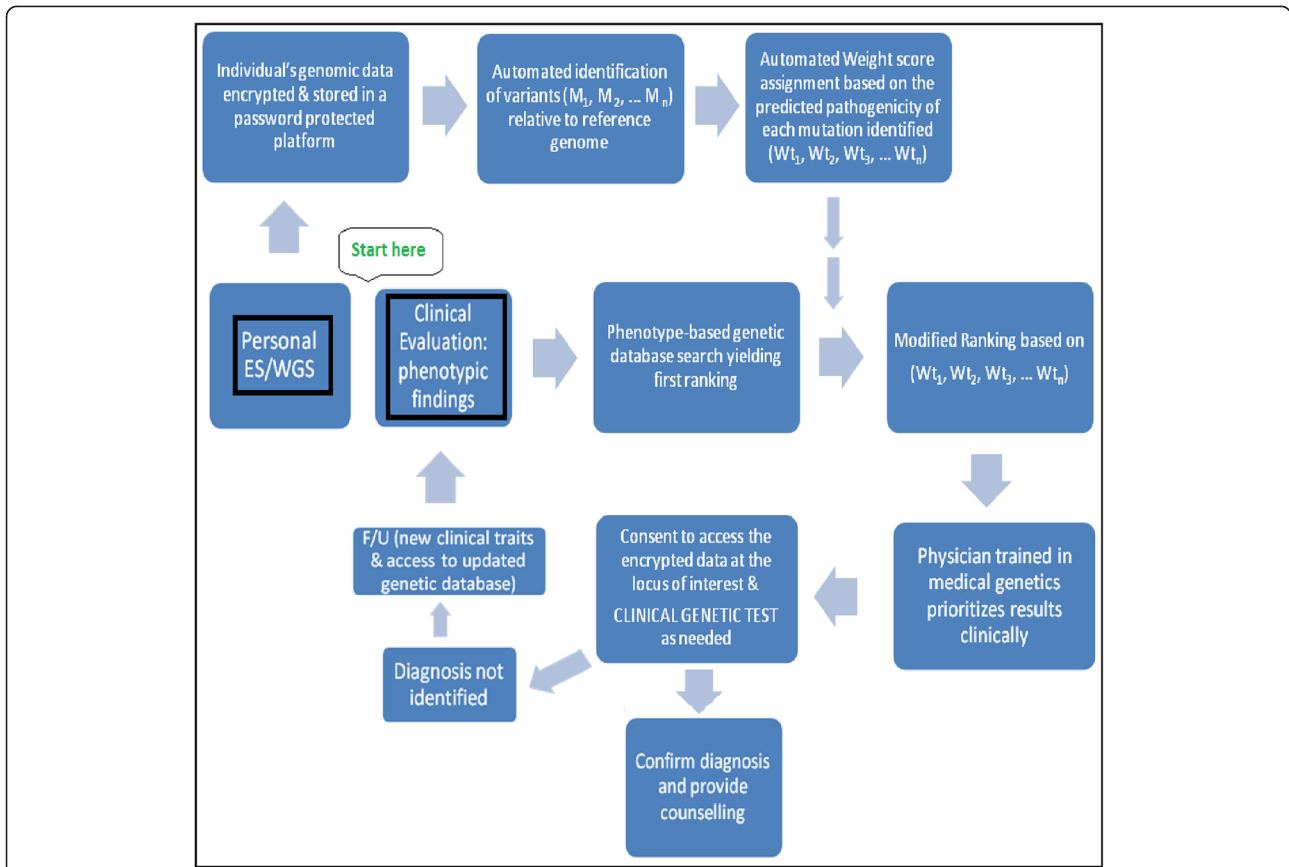
for gene discovery in rare monogenic syndromes [for instance, [28-31]]. As the cost of sequencing drops, the use of WGS is also becoming a research option [32-34]. The genes underlying most unsolved mendelian disorders known will likely soon be elucidated thanks to ES/WGS.

#### **Individualized Mutation-weighted Phenotype Search**

This manuscript proposes a new approach, Individualized Mutation-weighted Phenotype On-line Search (I-MPOS), which can help properly integrate genomic technologies in everyday clinical practice in the immediate future. In the illustrated approach (Figure. 1), patients have their exome/genome sequenced. The sequencing results are stored in an encrypted form on a platform (such as a chip, a memory key, or stored on a cloud) and can be accessed via a password only available to the individual patient. After obtaining authorization, the clinician can temporarily and anonymously upload the patient's encrypted data to a search engine, namely, Individualized Mutation-weighted Phenotype On-line Search Engine (I-MPOSE). I-MPOSE simultaneously operates on the patient's encrypted data and on a regularly updated database containing selected genetic diagnoses. Only the diagnoses for which there exists reliable evidence for a pathogenic relationship with at least one confirmed relevant region in the human genome and a well-described, more or less specific, subset of phenotypic characteristics are included in the database.

I-MPOSE identifies the genetic changes present in the patient's sequenced encrypted genome relative to the reference genome and assigns a weight score to each of them based on predetermined criteria integrated in the search algorithm. The pre-set criteria evolve as new knowledge is acquired and are equivalent, at any time, to the criteria used in the clinical evaluation of variants identified after a clinical genetic test (please refer to Figure. 2 for some examples based on current practice). The overall impact score calculated for each genetic variant identified corresponds to the level of certainty for its pathogenicity. The implementation of this approach can take advantage of the numerous databases and tools currently used in the interpretation of genetic variants generated by clinical genetic testing or ES/WGS clinical research (Figure. 3).

For this method to be most useful, the database has to include all known diseases which have been both phenotypically and genetically well characterized, at any given time, and be regularly updated based on research findings and clinical experience. Searching this database while taking into consideration both the patient's phenotype and his/her entire genomic data when ranking the results could aid enormously in the medical genetic evaluation of patients (please refer to discussion below).



**Figure 1 Schematic overview of I-MPOS, the new clinical genetic approach proposed.** Patients have their exome/genome sequenced and their encrypted data stored on a password-protected platform which remains at the disposal of the individual patient. A patient presents to clinic with a specific medical concern. The physician performs a clinical evaluation and identifies some important features (“phenotype-first” approach). He or she then performs a database search using keywords related to the clinically assessed phenotype, as presently done, thereby providing an initial ranking of possible genetic diseases. This initial ranking is then adjusted by I-MPOSE based on the weight scores automatically assigned to the mutations identified by ES/WGS in the patient’s genes/loci known to be linked to each genetic disease; thus providing a second ranking of the possible genetic diseases. I-MPOSE simultaneously operates on the patient’s encrypted data and on a regularly updated database containing all well characterized genetic diagnoses. It is run during every clinical visit so as to incorporate new findings from clinical evaluation, as well as, new genetic knowledge incorporated in the regularly updated database.

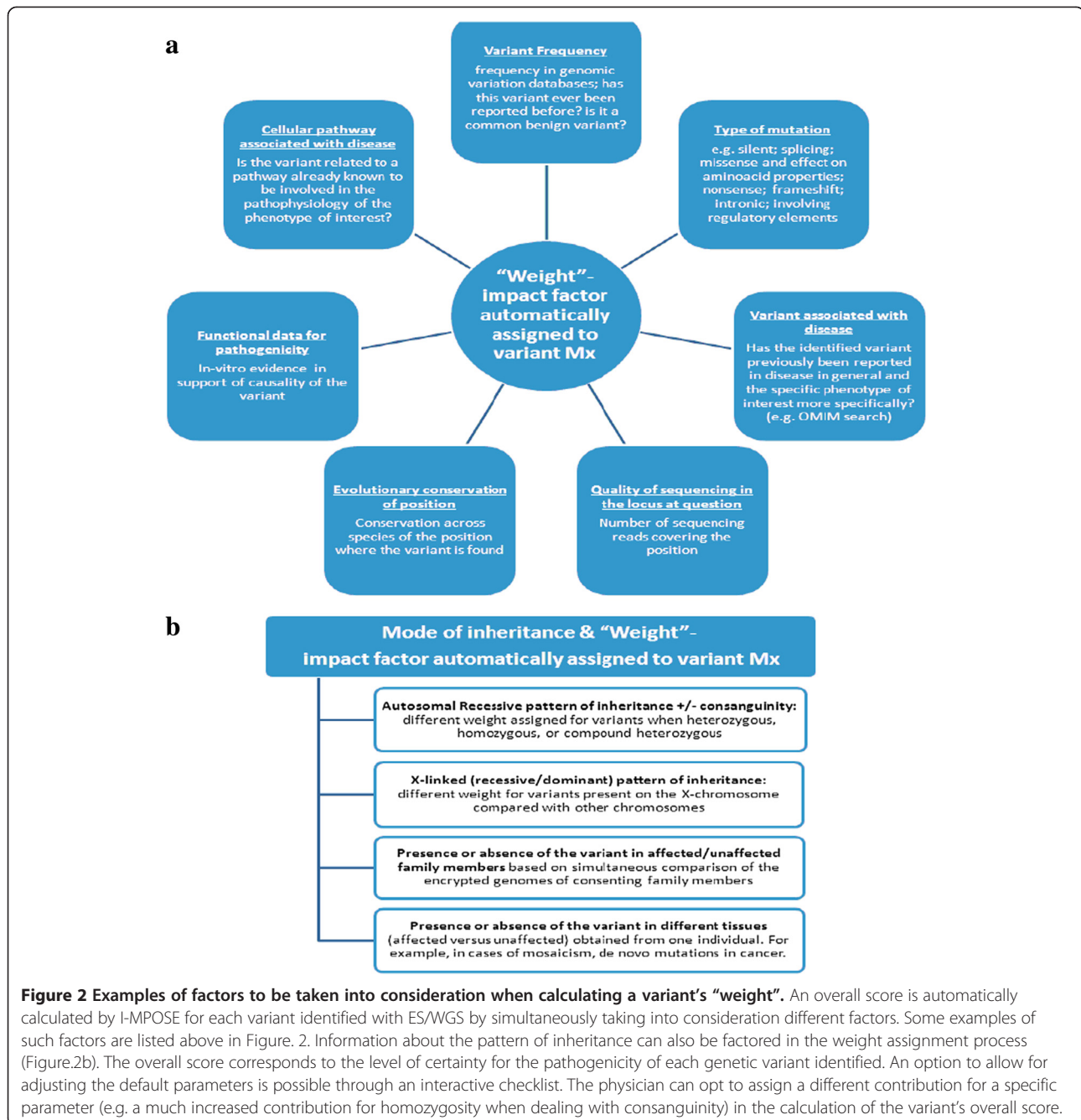
In summary, the patient presents to clinic with a specific medical concern. As currently practiced, the physician performs a clinical evaluation and identifies some important features (“phenotype-first” approach). He or she then performs a database search using keywords related to the clinically assessed phenotype, as presently done, thereby providing an initial ranking of possible genetic diseases. This initial ranking of possible genetic diseases is then adjusted by I-MPOSE based on the assigned weight scores for the variants identified by exome/genome sequencing in this patient; thus providing a second ranking of possible genetic diseases based on the said adjustment.

Given that the patient’s personal genome data is encrypted, the physician does not yet know anything about the exact mutations identified by genome/exome sequencing of the patient evaluated. However, after

viewing the adjusted ranking results, the physician could provide counselling to the patient and obtain informed consent to have access to the genomic data at the specific locus of interest before sending a sample for clinical genetic testing to confirm the diagnosis.

#### EXAMPLE

A patient P referred to genetics for hypotonia suffers from Smith Lemli Opitz (SLO) syndrome. The patient P’s exome has been sequenced. According to the crude data from exome sequencing, patient P has a homozygous mutation, in the *DHCR7* gene, known to cause SLO. The patient is clinically evaluated by a medical geneticist following exome sequencing. The geneticist is not aware of the mutation in the *DHCR7* gene nor that the patient has SLO syndrome. The only positive indications the geneticist identifies as important during the



evaluation are “hypotonia” and a “heart defect”. The geneticist performs two searches, a standard search as well as a search with I-MPOSE. The standard search is performed on the previously-mentioned website, OMIM [<http://omim.org/>].

The OMIM search results for the keywords “hypotonia AND heart defect” garnered 175 hits. SLO based on the search terms had a match score (Y) and was ranked #23 on this list. The I-MPOSE search was also performed on the OMIM website in tandem in conjunction with the patient’s encrypted sequencing data. The geneticist did

not look at the ES data per se. He/she simply ran during the search in tandem with the clinical criteria “hypotonia AND heart defect” the ranking system previously described, which took into consideration the automatically assigned impact score of each variant identified by ES. Given that the identified pathogenic mutation in *DHCR7* was assigned by I-MPOSE a high impact score (e.g. X), this time the cumulative score of SLO (Y x X) was relatively higher than that of most other diseases on the list. Hence, SLO was ranked higher on the search results list and the geneticist entertained the diagnosis of

## Examples of existing Databases

- **Online Mendelian Inheritance in Man (OMIM):** A regularly updated catalogue of human mendelian disease (<http://www.omim.org>);
- **Syndrome to Gene (S2G):** A candidate gene prediction tool; S2G includes a phenotype OMIM-based search engine used to prioritize candidate genes based on their association with genes known to underlie similar clinical syndromes (<http://fohs.bgu.ac.il/s2g/index.html>);
- **MutaDATABASE:** A centralized & standardized DNA variation database with clinical features of patients with monogenic disease (<http://www.mutadatabase.org/>);
- **The 1000 genome database:** A collection of all the variations in healthy individuals (<http://www.1000genomes.org/>);
- **The Single Nucleotide Polymorphism database (dbSNP)**  
<http://www.ncbi.nlm.nih.gov/projects/SNP/>
- **Locus Specific Mutation Databases** (<http://www.hgvs.org/dblist/glsdb.html>);
- **Human Genome Variation** (<http://www.hgvs.org/>);
- **Human Gene Mutation Database** (<http://www.hgmd.cf.ac.uk/ac/>);
- **Genotype to Phenotype databases GEN2PHEN** ([www.gen2phen.org](http://www.gen2phen.org))

## Variants management & analysis tools; Examples

- **Burrows-Wheeler Aligner (BWA)** (<http://bio-bwa.sourceforge.net/>)
- **Sequence Variant Analyzer (SVA)**, <http://www.svaproject.org/>
- **Sequence Alignment/Map tools (SAMTools)**  
(<http://samtools.sourceforge.net/>)
- **The Genome Analysis Toolkit – GSA (GATK)**  
<ftp://ftp.broadinstitute.org/pub/gsa/GenomeAnalysisTK/GenomeAnalysisTK-latest.tar.bz2>
- **Galaxy computational framework** (<http://usegalaxy.org>)
- **Database for Annotation, Visualization and Integration Discovery (DAVID)**  
• (<http://david.niaid.nih.gov>)
- **ANNOVAR** Functional annotation of genetic variants,  
(<http://www.openbioinformatics.org/annovar/>)
- **Screening of Non-Acceptable Polymorphisms (SNAP)**  
<http://www.ngsl.org.uk/Manchester/page/snap-screening-nonacceptable-polymorphisms>
- **Variant Annotation, Analysis and Search tool (VAAST)** (<http://www.yandell-lab.org/software/vaast.html>)
- **Mutation@A Glance** (<http://rapid.rcai.riken.jp/mutation/>)
- **Sorting Intolerant From Tolerant (SIFT)** (<http://sift.jcvi.org/>)
- **Polymorphism Phenotyping (PolyPhen)**  
(<http://genetics.bwh.harvard.edu/pph/>);
- **Mutation assessor** (<http://mutationassessor.org/>)
- **SNP Effect Predictor** (<http://snpeff.sourceforge.net/>)
- **University of California, Santa Cruz (UCSC) Genome Browser, for conservation across species** <http://genome.ucsc.edu/>
- **SNPtoGO characterizing SNPs by enriched GO terms**  
(<https://webtools.imbs.uni-luebeck.de/snptogo>)

**Figure 3** Examples of ongoing projects & tools illustrating that the necessary infrastructure for I-MPOSE is already available. Some examples of existing databases for human disease and genetic variation, as well as, tools available for the annotation of variants identified by ES/WGS.

SLO. The physician decided to send for *DHCR7* gene clinical testing to confirm the diagnosis. In this example, as per current practice, before undertaking clinical molecular testing the geneticist could have chosen to send a less-expensive non-genetic lab test named a sterol profile, which if positive would have further increased the physician's a priori suspicion for SLO. Similarly, the physician, after discussing his suspicion of SLO syndrome with the patient (or the family of the patient as appropriate), could have obtained consent to specifically analyze the patient's encrypted exome sequencing data in the *DHCR7* gene locus. This analysis could be

performed in collaboration with the clinical molecular lab director before ultimately pursuing the clinical diagnostic test. A more careful analysis of the changes identified by ES/WGS in this locus would have thus provided further evidence to support that the diagnosis of SLO syndrome needed to be entertained.

### Clinical Implications

Conventional approaches often require multiple genetic tests before a molecular diagnosis is reached and each clinical genetic test is often very expensive. The proposed approach, which follows the existing "phenotype-first"

medical model, allows for better prioritization of the genes to be tested in a clinical lab and is overall efficient from a resource, time and personnel point of view both in the clinic and the laboratory.

Based on the lessons learned from the past [35,36], an evidence-based, cautious approach to genomic medicine will be more efficient than premature efforts to integrate these technologies into clinical practice [5]. The I-MPOS approach allows ES/WGS to be used as a screening test rather than a clinical diagnostic test. A confirmatory clinical diagnostic test is still required, thus at least temporarily addressing the issue of patented genes, while ensuring reliable test results. The object of this method is to increase the diagnostic yield of a clinical evaluation in a cost-effective fashion and to decrease the time to diagnosis. For the remaining unresolved cases, the clinician needs to resort back to conventional approaches, which reduces false negatives, as well as, accounts for other known genetic explanations (e.g. epigenetic changes) not accounted for by ES/WGS. The clinician remains responsible for identifying and clinically prioritizing testing for the clinically suspected syndromes using the whole spectrum of clinical genetic testing modalities available. I-MPOSE is, simply, an additional tool at the clinician's disposal to make the foregoing assessment. Its limitations are minimized as long as qualified medical geneticists perform the evaluation, at least in the early stages of medical genomics, until other health professionals are adequately educated regarding its value, limitations, and appropriate clinical use.

In practice, the clinician does not have access to the patient's genome/exome sequencing results. The results remain encrypted and password-protected so that the patient retains full control over his/her data and how it is utilised. The patient only needs to consent that the clinician can explore the genetic causes which can potentially explain the specific phenotype/medical-issue at hand. The clinical terms entered in a search are selected based on the diseases of interest for which the family is seeking medical advice. Patients whose diagnosis has not been made will be seen in follow-up, as per present practice, and I-MPOSE will be run during every clinical visit so as to incorporate new findings from clinical evaluation, as well as, new genetic knowledge incorporated in the regularly updated database (Figure.1 for diagram). Hence, I-MPOS addresses many of the challenges associated with integrating genomic technologies into clinical practice (see Table 2).

Some individuals are likely to want to learn more from their genomic data as soon as ES/WGS is performed and this could be available in private, if properly regulated. However, in most centers, investing financial resources for such services would place pressure on an already strained health-care system [37,38]. Until the

**Table 2 Benefits of the I-MPOS approach**

---

-Simplification of pre-test counselling, informed consent and post-test counselling
-Protection of the patient's privacy and autonomy
-Genomic sequencing, even during newborn period, practical, ethical and clinically useful
-Increased diagnostic yield; cost-effective approach; decreased time to diagnosis
-Overcomes challenges such as shortage of personnel, the management of the huge amount of data generated, incidental findings, VUS, and duty to re-contact.
-Incremental integration of genomic technologies into clinical practice.
-Standardized approach in medical genomics, even where resources are limited
-Promotion of genomic research
-Refinement of clinical phenotype: partial matches for known syndromes; biochemical phenotypes
-Prevents over-medicalization of genomic data while enabling serving the patient's concerns
-Progressive education of both the health care personnel and the general public regarding personalized/preventive medicine

---

organisational, ethical, legal, and social issues are properly addressed, and the clinical validity/utility, as well as, the cost-effectiveness of such findings are proven, the clinician's access to the raw data will be restricted to rare challenging cases in a research setting. Hence, the I-MPOS paradigm helps to standardize care in the field of medical genomics based on realistic expectations which evolve as new knowledge is acquired, thus progressively increasing the quality of care provided to patients. Initially benefits will mostly focus on exome sequencing and rare mendelian conditions frequently encountered by clinical geneticists. The widespread utilization of ES/WGS will subsequently provide us lessons about more complex diseases.

Databases with patient-specific genomic data can store information of patients who decide to have their data stored for future reference or research. These databases will be managed similarly to the guidelines for "biobanks". With this approach, anonymized exome sequencing data from a great number of consenting individuals will likely be available and help us advance our knowledge of clinical application of genomic technologies. At a clinical level, on top of enabling earlier diagnosis [39], I-MPOS will aid in the refinement of the phenotype of known syndromes. For example, in conventional approaches only patients who fit very closely all or almost all characteristics of a described genetic syndrome are tested for the genes involved. The proposed approach provides for the partial matches of patients to be identified and their characteristics to enrich and refine the spectrum of clinical characteristics of any given syndrome.

### Existing infrastructure and Future Directions

The most central components needed for I-MPOS to be successfully implemented are the availability of well curated databases and the reliability of the “impact score” assignment for each variant identified. To aid in the latter, multiple computational tools exist (some of which are listed in Figure.3). There are also software tools which enable the user to take advantage of a gene filter (e.g. based on quality score, pathway, gene-ontology, OMIM information) and different biological databases when performing the functional annotation. One such example is the Sequence Variant Analyzer (SVA), which is also allowing users to explore the strength of the associations of identified variants with studied traits [40].

As previously stated, computational predictions about the pathogenicity of variants are not always reliable. However, combining the results of different such methods, as well as population and/or family-based data about the cosegregation of the variant with disease, using different statistical methods can be helpful in the calculation of an “overall impact-score” which is more reliable. For instance, Bayesian analysis can be used to factor in results of different computational tools in the overall score. Moreover, relative risk analysis and/or Transmission Disequilibrium Test (TDT) analysis for each variant at a population or family level can also be very useful in predicting a variant’s pathogenicity, particularly when ES/WGS usage becomes wide-spread [41-44]. Freely available statistical software like Family Based Association Test, a.k.a. FBAT, can be useful to this end [45]. In the future, the combination of different bioinformatics and statistical tools will be incrementally more important in the calculation of the “impact score” of each variant. It is also possible that metabolomics will lead to the identification of specific metabolic profiles/signatures associated with specific variants and that the (degree of) biochemical imbalance will constitute an important factor in predicting the pathogenicity, and thus in calculating the variant’s “impact factor”. There are already projects in metabolomics which in the future could be useful to this end [46-48].

In the first steps of the I-MPOS endeavor, though, one needs to be particularly cautious and focus on variants known to be pathogenic based on clinical experience. Already existing guidelines for variant interpretation can be used to classify each variant identified in one of four or six different groups with a different “impact factor” potentially assigned to each group [49,50]. Alternatively, initially the data of already existing databases listing annotated variants and classifying them as pathogenic or not can be the only factors included in the score assignment. Worldwide, ongoing efforts already aim to create comprehensive collections of validated associated variants, as

illustrated in Figure.3. Potentially useful curated databases include the Human Gene Mutation Database (HGMD) [51], Diagnostic Mutation Database (DMuDB) [52], Mutation Database [53], and the ClinVar database [54,55]. The implementation of common methods and clinical standards for data collection and reporting are of great importance for the optimal operation of I-MPOSE. The interpretation of genomic information will change over time as new risks are identified and others are refuted. Hence, the database to be used will need to be updated regularly with all the genetic changes confirmed based on research and clinical experience to cause or predispose to a genetic disease/phenotype. With regards to phenotypic data, an interface with freely available or private databases (such as OMIM, GeneReviews, London Medical Databases, or Possum Web) will be needed [56-59]. In order for the search to be optimized, a standardized vocabulary of phenotypic abnormalities encountered in human disease and the semantic relationships between them will be important. The Human Phenotype Ontology (HPO) and Phenomizer could be very useful to this end [60].

In the future, it would be useful if the database used for I-MPOSE included *genes* sorted into “priority tiers” based on the evidence of involvement in each genetic disease (“informatics disease panels” [61]). This could be expanded to include prioritization of genes involved in specific sub-phenotypes (rather than the genetic disease as an entity) like “microcephaly”, “short stature” or even more complex phenotypes such as biochemical profiles. In brief, illnesses, like schizophrenia or Type 1 Diabetes [62-66], with high heritability but complex inheritance, may be shown in the future to constitute a constellation of different pathogenetic processes leading to a similar phenotype. Some of these pathogenetic processes may ultimately have a specific signature not only at the level of the genome, but also at the level of the epigenome, transcriptome or metabolome. If the known such signature profiles are listed in the databases used by I-MPOSE, the data generated at the time of the clinical evaluation from RNA sequencing [67-69] and/or metabolomic profiling [70,71] or even epigenomic changes [72] over time in such a (symptomatic) patient could serve as a phenotypic trait refining the search.

The concept of “informatics disease panels” can also help in the identification of novel genes contributing to known genetic diseases/phenotypes. One can use different methods (see [73-82] for examples) to select candidate genes for each genetic disease/phenotype and also include them in the database, albeit with a lower “priority”, representing the amount of evidence/probability of the gene’s involvement in the respective genetic disease. For any given patient, when a phenotype is ranked highly after I-MPOSE search but no mutations are found in all genes documented to cause the disease, consent to



explore the *pre-selected* candidate genes for the phenotype at hand for mutations could be obtained.

In conclusion, the widespread utilization of ES/WGS will subsequently provide us lessons about complex genetic diseases. The weight factors and the assigning process will evolve based on emerging knowledge (e.g. to account for gene-gene or gene-environment interactions in complex diseases [83-89]). As research and genetic knowledge advance, the benefits of this approach will increase exponentially. I-MPOS could ultimately help cautiously drive the paradigm shift to personalized healthcare and eventually preventive medicine.

## Summary

This is an exciting time for medicine: the era of genomic medicine. However, the implementation of ES/WGS in clinical practice should only take place when the challenges posed by ethics and policy issues have been addressed. Unfortunately, it is very likely that the fast pace of development of these technologies and their ability to facilitate diagnosis, to reduce the incidence of deleterious genetic disorders, and to inform therapy will lead to their premature introduction in clinic [90]. As previously discussed, the era of WGS is just around the corner [91]. Already ES is employed by some direct-to-consumer genetic testing companies [19,92]. We now need to prepare the framework required for responsible and successful integration of ES/WGS in patient care; thus ensuring the appropriate and effective usage of genomic information [91]. It is widely recognized that automated systems and clear guidelines will be necessary in this process. Our current approaches need to be reassessed but not changed solely based on the availability of a new technology. For instance, currently genetic risk assessment for individuals who are not at increased risk based on family history or clinical presentation is generally restricted to newborn screening programs meeting specific criteria [5,93]. Is there adequate evidence to suggest that a departure from these standards will be beneficial to society? It appears that medical genetics is shifting from the present “phenotype-first” medical model to a “data-first” model which leads to multiple complexities. This manuscript discusses a “phenotype-first” approach, namely, Individualized Mutation-weighted Phenotype Search (I-MPOS), which could potentially render widespread use of ES/WGS, in the immediate future, practical, ethical and clinically useful. The proposed approach allows for a more efficient prioritization of the genes to be tested in a clinical lab based on both the patient’s phenotype and his/her entire genomic data. I-MPOS protects the patient’s privacy and autonomy and enables an incremental integration of genomic technologies into clinical practice. It allows for progressive education of both the health care personnel,

including medical geneticists and clinical laboratory scientists, and the general public regarding personalized/preventive medicine. The I-MPOS approach ensures that we do not over-medicalize genomic data but rather continue our current medical model which is based on serving the patient’s concerns. Service should not be solely driven by technology but rather by the medical needs and the extent to which a technology can be safely and effectively utilized.

The era of genomic medicine is already a reality.

## Abbreviations

ES: Exome Sequencing; WGS: Whole Genome Sequencing; I-MPOS: Individualized Mutation-weighted Phenotype On-line Search; I-MPOSE: Individualized Mutation-weighted Phenotype On-line Search Engine.

## Competing interests

Patent application in process: my intention is to ensure that the I-MPOS paradigm can be properly implemented in a timely fashion. As long as the necessary academic collaborations for the programme are available, non-commercial applications of this method will remain free of charge.

Received: 15 December 2011 Accepted: 30 June 2012

Published: 20 July 2012

## References

1. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, *et al*: Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010, **327**(5961):78–81.
2. Check Hayden E: Genome sequencing: the third generation. *Nature* 2009, **457**(7231):768–769.
3. Hunter DJ, Khoury MJ, Drazen JM: Letting the genome out of the bottle—will we get our wish? *N Engl J Med* 2008, **358**(2):105–107.
4. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J: Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011, **12**(11):745–755.
5. Sharp RR: Downsizing genomic medicine: approaching the ethical complexity of whole-genome sequencing by starting small. *Genet Med* 2011, **13**(3):191–194.
6. Ormond KE, Wheeler MT, Hudgins L, Klein TE, Butte AJ, Altman RB, Ashley EA, Greely HT: Challenges in the clinical application of whole-genome sequencing. *Lancet* 2010, **375**(9727):1749–1751.
7. Elias S, Annas GJ: Generic consent for genetic screening. *N Engl J Med* 1994, **330**(22):1611–1613.
8. McGuire AL, Caulfield T, Cho MK: Research ethics and the challenge of whole-genome sequencing. *Nat Rev Genet* 2008, **9**(2):152–156.
9. : A map of human genome variation from population-scale sequencing. *Nature* 2010, **467**(7319):1061–1073.
10. Ware JH: The limitations of risk factors as prognostic tools. *N Engl J Med* 2006, **355**(25):2615–2617.
11. Wald NJ, Hackshaw AK, Frost CD: When can a risk factor be used as a worthwhile screening test? *BMJ* 1999, **319**(7224):1562–1565.
12. Kaye J, Boddington P, de Vries J, Hawkins N, Melham K: Ethical implications of the use of whole genome methods in medical research. *Eur J Hum Genet* 2010, **18**(4):398–403.
13. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, *et al*: The diploid genome sequence of an individual human. *PLoS Biol* 2007, **5**(10):e254.
14. Guttmacher AE, Porteous ME, McInerney JD: Educating health-care professionals about genetics and genomics. *Nat Rev Genet* 2007, **8**(2):151–157.
15. Baars MJ, Henneman L, Ten Kate LP: Deficiency of knowledge of genetics and genetic tests among general practitioners, gynecologists, and pediatricians: a global problem. *Genet Med* 2005, **7**(9):605–610.

16. Kohane IS, Mandl KD, Taylor PL, Holm IA, Nigrin DJ, Kunkel LM: **Medicine. Reestablishing the researcher-patient compact.** *Science* 2007, **316**(5826):836–837.
17. Beaudet AL: **Ethical issues raised by common copy number variants and single nucleotide polymorphisms of certain and uncertain significance in general medical practice.** *Genome Med* 2010, **2**(7):42.
18. Deyo RA: **Cascade effects of medical technology.** *Annu Rev Public Health* 2002, **23**:23–44.
19. McGuire AL, Burke W: **An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons.** *JAMA* 2008, **300**(22):2669–2671.
20. Snyder M, Du J, Gerstein M: **Personal genome sequencing: current approaches and challenges.** *Genes Dev* 2010, **24**(5):423–431.
21. Jensen K, Murray F: **Intellectual property. Enhanced: intellectual property landscape of the human genome.** *Science* 2005, **310**(5746):239–240.
22. Kesselheim AS, Karlawish J: **Biomarkers Unbound - The Supreme Court's Ruling on Diagnostic-Test Patents.** *N Engl J Med* 2012, **366**(25):2338–40.
23. Huys I, Matthijs G, Van Overwalle G: **The fate and future of patents on human genes and genetic diagnostic methods.** *Nat Rev Genet* 2012, **13**(6):441–448.
24. Guttmacher AE, McGuire AL, Ponder B, Stefansson K: **Personalized genomic information: preparing for the future of genetic medicine.** *Nat Rev Genet* 2010, **11**(2):161–165.
25. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE: **ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007.** *Genet Med* 2008, **10**(4):294–300.
26. Dorfman R, Nalpathamkalam T, Taylor C, Gonska T, Keenan K, Yuan XW, Corey M, Tsui LC, Zielenski J, Durie P: **Do common in silico tools predict the clinical consequences of amino-acid substitutions in the CFTR gene?** *Clin Genet* 2010, **77**(5):464–473.
27. Thusberg J, Olatubosun A, Vihinen M: **Performance of mutation pathogenicity prediction methods on missense variants.** *Hum Mutat* 2011, **32**(4):358–368.
28. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: **Exome sequencing identifies the cause of a mendelian disorder.** *Nat Genet* 2010, **42**(1):30–35.
29. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, et al: **Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.** *Nat Genet* 2010, **42**(9):790–793.
30. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, et al: **De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.** *Nat Genet* 2010, **42**(6):483–485.
31. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Rieuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG: **Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome.** *Am J Hum Genet* 2010, **87**(3):418–423.
32. Green ED, Guyer MS: **Charting a course for genomic medicine from base pairs to bedside.** *Nature* 2011, **470**(7333):204–213.
33. Rios J, Stein E, Shendure J, Hobbs HH, Cohen JC: **Identification by whole-genome resequencing of gene defect responsible for severe hypercholesterolemia.** *Hum Mol Genet* 2010, **19**(22):4313–4318.
34. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irlkat RK, Tom W, Fantin NJ, et al: **Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy.** *N Engl J Med* 2010, **362**(13):1181–1191.
35. Kevles DJ: **From eugenics to patents: genetics, law, and human rights.** *Ann Hum Genet* 2011, **75**(3):326–333.
36. Cavazzana-Calvo M, Thrasher A, Mavilio F: **The future of gene therapy.** *Nature* 2004, **427**(6977):779–781.
37. McGuire AL, Cho MK, McGuire SE, Caulfield T: **Medicine. The future of personal genomics.** *Science* 2007, **317**(5845):1687.
38. Burke W, Psaty BM: **Personalized medicine in the era of genomics.** *JAMA* 2007, **298**(14):1682–1684.
39. Jimenez-Escrig A, Gobernado I, Garcia-Villanueva M, Sanchez-Herranz A: **Autosomal recessive Emery-Dreifuss muscular dystrophy caused by a novel mutation (R225Q) in the lamin A/C gene identified by exome sequencing.** *Muscle Nerve* 2012, **45**(4):605–610.
40. Ge D, Ruzzo EK, Shianna KV, He M, Pelak K, Heinzen EL, Need AC, Cirulli ET, Maia JM, Dickson SP, et al: **SVA: software for annotating and visualizing sequenced human genomes.** *Bioinformatics* 2011, **27**(14):1998–2000.
41. Londono D, Buyske S, Finch SJ, Sharma S, Wise CA, Gordon D: **TDT-HET: a new transmission disequilibrium test that incorporates locus heterogeneity into the analysis of family-based association data.** *BMC Bioinforma* 2012, **13**:13.
42. Ott J, Kamatani Y, Lathrop M: **Family-based designs for genome-wide association studies.** *Nat Rev Genet* 2011, **12**(7):465–474.
43. Zhu Y, Xiong M: **Family-based association studies for next-generation sequencing.** *Am J Hum Genet* 2012, **90**(6):1028–1045.
44. Fang S, Sha Q, Zhang S: **Two adaptive weighting methods to test for rare variant associations in family-based designs.** *Genet Epidemiol* 2012, **36**(5):499–507.
45. Laird NM, Horvath S, Xu X: **Implementing a unified approach to family-based tests of association.** *Genet Epidemiol* 2000, **19**(Suppl 1):S36–42.
46. Go EP: **Database resources in metabolomics: an overview.** *J Neuroimmune Pharmacol* 2010, **5**(1):18–30.
47. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, et al: **The human serum metabolome.** *PLoS One* 2011, **6**(2):e16957.
48. Huang H, Hu ZZ, Arighi CN, Wu CH: **Integration of bioinformatics resources for functional analysis of gene expression and proteomic data.** *Front Biosci* 2007, **12**:5071–5088.
49. Bell J, Bodmer D, Sistermans E, Ramsden SC: **Practice guidelines for the interpretation and reporting of unclassified variants (UVs) in clinical molecular genetics.** UK: Guidelines ratified by the UK CMGS (11th January, 2008) and the VGKL (22nd October, 2007); 2007. [http://cmgsweb.shared.hosting.zen.co.uk/BPGs/Best\\_practice\\_Guidelines.htm](http://cmgsweb.shared.hosting.zen.co.uk/BPGs/Best_practice_Guidelines.htm).
50. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE: **ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007.** *Genet Med* 2008, **10**(4):294–300.
51. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN: **The Human Gene Mutation Database: 2008 update.** *Genome Med* 2009, **1**(1):13.
52. *The Diagnostic Mutation Database. (DMuDB)* [<http://www.ngsl.org.uk/Manchester/projects/informatics/dmudb>]
53. Bale S, Devisscher M, Van Criekinge W, Rehm HL, Decouttere F, Nussbaum R, Dunnen JT, Willems P: **MutaDATABASE: a centralized and standardized DNA variation database.** *Nat Biotechnol* 2011, **29**(2):117–118.
54. *ClinVar.* <http://www.ncbi.nlm.nih.gov/clinvar>.
55. Rehm H, Martin C, Bale S, Hegde M, Willems P, Miller D, Lyon E, Ferber M, Das S, Riggs E, Kaminsky E, Faucett W, Paschall J, Maglott D, Ostell J, Ledbetter D, Nussbaum R: *ClinVar Community Group Centralizing the Deposition and Curation of Human Mutations.*; [abstract] <http://submissions.miracdc.com/acmg/ContentInfo.aspx?conID=2599>.
56. Amberger J, Bocchini C, Hamosh A: **A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)).** *Hum Mutat* 2011, **32**(5):564–567.
57. *GeneReviews™.* [<http://www.ncbi.nlm.nih.gov/books/NBK1116/>]
58. *POSSUM Web.* [<http://www.possum.net.au/>]
59. *The London Medical Databases.* [<http://www.lmdatabases.com/>]
60. *The Human Phenotype Ontology.* [<http://www.human-phenotype-ontology.org/>]
61. Crooks K, Berg J, Booker J, Evans J, Foreman K, Weck K: **Separating the wheat from the chaff: A gene-based approach for prioritizing analysis of genetic variants in diagnostic whole-exome sequencing in NCGENES.** 2012. [abstract] <http://submissions.miracdc.com/acmg/ContentInfo.aspx?conID=2855>.
62. Brazo P, Dollfus S: **[Syndromic and diagnostic heterogeneity of schizophrenia].** *Encephale* 1997, **23 Spec No 2**:20–24.
63. Carpenter WT Jr: **Schizophrenia: disease, syndrome, or dimensions?** *Fam Process* 2007, **46**(2):199–206.
64. Sullivan PF: **The genetics of schizophrenia.** *PLoS Med* 2005, **2**(7):e212.
65. Greenwood TA, Light GA, Swerdlow NR, Radant AD, Braff DL: **Association analysis of 94 candidate genes and schizophrenia-related endophenotypes.** *PLoS One* 2012, **7**(1):e29630.

66. Kyvik KO, Green A, Beck-Nielsen H: **Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins.** *BMJ* 1995, **311**(7010):913–917.
67. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Phillipakis AA, del Angel G, Rivas MA, Hanna M, et al: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nat Genet* 2011, **43**(5):491–498.
68. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK: **Understanding mechanisms underlying human gene expression variation with RNA sequencing.** *Nature* 2010, **464**(7289):768–772.
69. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET: **Transcriptome genetics using second generation sequencing in a Caucasian population.** *Nature* 2010, **464**(7289):773–777.
70. Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, Schadt EE: **Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation.** *PLoS Biol* 2012, **10**(4):e1001301.
71. Snyder M, Weissman S, Gerstein M: **Personal phenotypes to go with personal genomes.** *Mol Syst Biol* 2009, **5**:273.
72. Maunakea AK, Chepelev I, Zhao K: **Epigenome mapping in normal and disease States.** *Circ Res* 2010, **107**(3):327–339.
73. Kanehisa M: **The KEGG database.** *Novartis Found Symp* 2002, **247**:91–101. discussion 101–103, 119–128, 244–152.
74. Epstein C, Erickson R, Wynshaw-Boris A: *Inborn Errors of Development.* Oxford: Oxford University Press; 2008.
75. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, et al: **The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.** *Nucleic Acids Res* 2010, **38**(Web Server issue):W214–220.
76. Raychaudhuri S, Plenge RM, Rossin EJ, Ng AC, Purcell SM, Sklar P, Scolnick EM, Xavier RJ, Altshuler D, Daly MJ: **Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions.** *PLoS Genet* 2009, **5**(6):e1000534.
77. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, Catanese JJ, Xie G, Stahl EA, Chen R, et al: **Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk.** *Nat Genet* 2009, **41**(12):1313–1318.
78. Jani SD, Argraves GL, Barth JL, Argraves WS: **GeneMesh: a web-based microarray analysis tool for relating differentially expressed genes to MeSH terms.** *BMC Bioinforma* 2010, **11**:166.
79. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, Zoghbi HY: **Protein interactome reveals converging molecular pathways among autism disorders.** *Sci Transl Med* 2011, **3**(86):86ra49.
80. Kou Y, Betancur C, Xu H, Buxbaum JD: **Ma'ayan A: Network- and attribute-based classifiers can prioritize genes and pathways for autism spectrum disorders and intellectual disability.** *Am J Med Genet C Semin Med Genet* 2012, **160C**(2):130–142.
81. Connor SC, Hansen MK, Corner A, Smith RF, Ryan TE: **Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes.** *Mol Biosyst* 2010, **6**(5):909–921.
82. Johnson JM, Yu T, Strobel FH, Jones DP: **A practical approach to detect unique metabolic patterns for personalized medicine.** *Analyst* 2010, **135**(11):2864–2870.
83. Cordell HJ: **Detecting gene-gene interactions that underlie human diseases.** *Nat Rev Genet* 2009, **10**(6):392–404.
84. Bureau A, Merette C, Croteau J, Fournier A, Chagnon YC, Roy MA, Maziade M: **A new strategy for linkage analysis under epistasis taking into account genetic heterogeneity.** *Hum Hered* 2009, **68**(4):231–242.
85. Lincoln MR, Ramagopalan SV, Chao MJ, Herrera BM, Deluca GC, Orton SM, Dymant DA, Sadovnick AD, Ebers GC: **Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility.** *Proc Natl Acad Sci U S A* 2009, **106**(18):7542–7547.
86. Gauderman WJ: **Candidate gene association analysis for a quantitative trait, using parent-offspring trios.** *Genet Epidemiol* 2003, **25**(4):327–338.
87. Hunter DJ: **Gene-environment interactions in human diseases.** *Nat Rev Genet* 2005, **6**(4):287–298.
88. Thomas D: **Gene-environment-wide association studies: emerging approaches.** *Nat Rev Genet* 2010, **11**(4):259–272.
89. Khoury MJ, Flanders WD: **Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls!** *Am J Epidemiol* 1996, **144**(3):207–213.
90. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, et al: **Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.** *Genet Med* 2011, **13**(3):255–262.
91. Feero WG, Guttmacher AE, Collins FS: **The genome gets personal—almost.** *JAMA* 2008, **299**(11):1351–1352.
92. Ikediobi ON: **Personalized medicine: are we there yet?** *Pharmacogenomics J* 2009, **9**(2):85.
93. Wilson JM, Jungner YG: **Principles and practice of mass screening for disease.** *Bol Oficina Sanit Panam* 1968, **65**(4):281–393.

doi:10.1186/1755-8794-5-31

**Cite this article as:** Trakadis: Patient-controlled encrypted genomic data: an approach to advance clinical genomics. *BMC Medical Genomics* 2012 5:31.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

