Evidence-based medicine and big genomic data

John P.A. Ioannidis[1],[*] and **Muin J. Khoury**[2]

[1]Stanford Prevention Research Center, Meta-Research Innovation Center at Stanford (METRICS), and Departments of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, Stanford University, Stanford, CA, USA

[2]Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract

Genomic and other related big data (Big Genomic Data, BGD for short) are ushering a new era of precision medicine. This overview discusses whether principles of evidence-based medicine hold true for BGD and how they should be operationalized in the current era. Major evidence-based medicine principles include the systematic identification, description and analysis of the validity and utility of BGD, the combination of individual clinical expertise with individual patient needs and preferences, and the focus on obtaining experimental evidence, whenever possible. BGD emphasize information of single patients with an overemphasis on N-of-1 trials to personalize treatment. However, large-scale comparative population data remain indispensable for meaningful translation of BGD personalized information. The impact of BGD on population health depends on its ability to affect large segments of the population. While several frameworks have been proposed to facilitate and standardize decision making for use of genomic tests, there are new caveats that arise from BGD that extend beyond the limitations that were applicable for more simple genetic tests. Non-evidence-based use of BGD may be harmful and result in major waste of healthcare resources. Randomized controlled trials will continue to be the strongest arbitrator for the clinical utility of genomic technologies, including BGD. Research on BGD needs to focus not only on finding robust predictive associations (clinical validity) but also more importantly on evaluating the balance of health benefits and potential harms (clinical utility), as well as implementation challenges. Appropriate features of such useful research on BGD are discussed.

Introduction

The emergence of genomic sequencing technologies and other-omic information (e.g. transcriptomics or metabolomics) along with a large amount of digital big data on individuals and populations are poised to usher in a new era of precision medicine (1). Very large numbers of individuals have or will soon have access to such 'big genomic data' (BGD). Concurrently, BGD may be coupled also to large-scale information for environmental exposures and lifestyle (2). People may use BGD for health-related reasons in the context of healthcare delivery or on their own initiative via direct-to-consumer offerings (3).

[*]To whom correspondence should be addressed at: Stanford Prevention Research Center, 1265 Welch Road, Medical School Office Building, Room X306, Stanford, CA 94305, USA. Tel: +1 6507045584; jioannid@stanford.edu.

Several questions arise: Should the principles of evidence-based medicine (EBM) change with BGD? Should old hierarchies of evidence be modified with personalized big data? What might be realistic expectations for precision health goals? What framework should guide decision making on the use of BGD? Do we need randomized controlled trials (RCTs) to assess the clinical utility of BGD? Finally, how do we design an evidence-based research agenda to maximize utility for BGD? This overview tries to address these questions.

Should the principles of evidence-based medicine change with big genomic data?

The term 'evidence-based' was introduced (4) to herald the need of '…consciously anchoring a policy, not to current practices or the beliefs of experts, but to experimental evidence… The pertinent evidence must be identified, described, and analyzed'. EBM is defined (5) as 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients… [It] means integrating individual clinical expertise with the best available external clinical evidence from systematic research'.

These principles are still relevant in the era of precision medicine. BGD should not be viewed as a shortcut for EBM (6). Identifying, describing and analyzing pertinent evidence is essential for BGD, as for any type of evidence. The challenge becomes greater as the mass of information increases, fragmented across multiple facilities, sites, healthcare systems, electronic health record (EHR) systems or personal files of billions of individuals. There may be a variable amount of sharing, and variable adoption of plans towards cumulative knowledge. The focus on 'making decisions about …individual patients' is more important than ever, as BGD aims to empower individuals in personalized decisions.

The definition of EBM also includes 'individual clinical expertise', while there is speculation that BGD may allow health decisions by individuals without involving any health practitioner. However, shared decision making involving both a patient and a clinician (7) or other specialized interlocutor is more realistic. It is currently unclear who is the best genomic interlocutor of the patient/individual (trained clinicians, genetic counselors and others) and how best to implement the use of genomic information. A systematic review of 283 implementation studies for genetic information concludes that there is very weak evidence and no solid conclusions (8). The complexity of interpreting genomic information (9) makes simplified solutions unrealistic. While the cost of genomic and other-omic technologies will continue to decrease, the cost for interpreting genomics and other big data may escalate.

Furthermore, the need for systematic evidence synthesis remains strong. Experimental evidence may be sparse/nonexistent (as discussed below) and one would then have to decide whether other types of evidence can fill the gap.

Should old hierarchies of evidence be modified?

A fundamental tension is whether BGD disrupt traditional hierarchies of evidence (10). Traditional hierarchies place systematic reviews and meta-analyses at the top, followed by randomized trials, then observational studies and case reports on single patients and expert opinion at the bottom. BGD often emphasize information of single patients under the rubric of personalized/precision medicine/health. Emphasis is placed on mechanistic insights from single cases (10). Moreover, there are high expectations for N-of-1 trials aiming to identify personalized treatments that are guided by BGD.

N-of-1 trials are actually an old concept. There was substantial interest in N-of-1 trials in the 1980s and 1990s (11,12) and some old hierarchies of evidence had even placed N-of-1 trials at the top of the evidence pyramid. Getting reliable information about how to treat individuals seemed more attractive than summary average results with unclear relevance for any individual. N-of-1 trials, however, have not gained much traction, because they had major caveats. Specifically, when sequentially testing different treatment options, insights obtained are unreliable when the disease lacks a steady natural history, when there are substantial carry-over effects of prior treatment options, when the treatment effect is influenced by previous treatment choices and their order, and when the disease has a fatal outcome and a relatively short course (13). The most successful applications of N-of-1 trials to-date (14) have been for chronic, lifetime diseases like osteo-arthritis, chronic neuropathic pain or attention-deficit hyper-activity disorders, where these caveats largely do not exist. Even for these conditions, the proportions of participants persisting with the joint patient-doctor decision 12 months after trial completion were modest (32%, 45% and 70% for the three conditions, respectively) (14). Conversely, most of the currently contemplated uses of N-of-1 approaches for treatment based on big data of individuals, in particular late-stage malignancies, suffer from many of the caveats that diminish the value and reliability of N-of-1 trials. Inferences may be unreliable in these settings. N-of-1 trials may become most useful, if enough big data can be collected to define an unequivocal phenotype and, concurrently, a perfectly tailored effective treatment can be found for this phenotype. At best, this can be seen as speculative, work in progress, as we discuss below. In the meanwhile, large-scale population-level evidence would continue to be indispensable to juxtapose individual patient profiles and experiences against such large-scale data. This evidence will be most reliable if it has been collected with accurate measurements (analytical validity), has shown reproducible associations and effects of interest across diverse populations (clinical validity) and has been scrutinized systematically to ensure clinical validity without major bias.

What might be realistic expectations for precision health goals?

According to Collins and Varmus (1) precision medicine aims 'to give everyone the best chance at good health'. Healthcare is customized, with medical decisions, practices or products tailored to the individual patient. However, EBM also focuses on individual patients (5), so is this a new concept? Perhaps the one difference is that EBM has accepted that, in order to get best evidence for an individual, one typically needs large-scale data on the greater population. Conversely, many proponents of precision medicine believe that

amassing large-scale data on a single person would suffice to get the best evidence for treating the person. This is a widespread misunderstanding (15).

Given that the individual and the population stand at the two ends of the spectrum, by definition, precision medicine, as advocated by many proponents, may have a tiny or negligible impact at a population level. For single patients, it may indeed have maximal, optimized benefits. However, this hypothetical scenario has to be proven on a case-by-case basis. Even if it were true, one would have to expect that many people could reap such major personalized benefits, in order to make the precision medicine/population health agenda worthwhile. The most common application of BGD to date is in cancer genomics. However, in the National Cancer Institute-Molecular Analysis for Therapy Choice (NCI-MATCH) trial (16) that screens cancer patients for tumor mutations, only 2.5% of patients match to a molecularly individualized treatment. This does not reflect just a difficulty to identify precise molecular profiles. Even when molecular profiles can be identified, most often there is yet no good treatment to match to them. Moreover, examples where new drugs have been approved based primarily on mechanistic knowledge without randomized trials become more common; however, they mostly pertain to conditions with negligible population burden (e.g. approval of ivacaftor for cystic fibrosis caused by 23 new residual function mutations based on the results of functional assays and approval of pembrolizumab for progressive metastatic solid tumors with microsatellite instability or mismatch repair deficiency) (17,18).

Similarly, there is ambiguity even about what big data mean (19). Big data typically carry minimal useful information content per unit, hence a big amount is needed to make them useful for practical purposes. The more insignificant the content of information per unit, the bigger the big data required to start having some utility. This is the exact opposite of what Bradford Hill, one of the fathers of modern epidemiology, would cherish. He felt confident when calculations were possible to repeat on the back of an envelope. For large effects, for example association of smoking with lung cancer, 2×2 tabulation and calculation of the huge odds ratio are doable on the back of an envelope. For complicated, perplexing big data, this is impossible. Big data black boxes typically do not inspire confidence. The ability to handle a complex, convoluted black box in everyday circumstances is unclear. It is even possible that people with greater access to big data eventually have worse outcomes. Some big data may confuse, overwhelm and mobilize people in vain to change their lives or seek treatment adventures for no good reason. A reliable framework is needed to decide which big data should be used.

What framework should guide decision making on the use of big genomic data?

The National Academies of Sciences, Engineering, and Medicine report on 'An Evidence Framework for Genetic Testing' (20) offers a consensus framework for decision making on genetic/genomic tests in clinical care. It builds on previous frameworks, including the United States Preventive Services Task Force approach for assessing preventive interventions (21); the Fryback–Thornbury hierarchy (22); the ACCE (Analytic Validity, Clinical Validity, Clinical Utility, and Ethical, Legal, and Social Implications) framework (23); the EGAPP

(Evaluation of Genomic Applications in Practice and Prevention) standards (24); the Genetic testing Evidence Tracking Tool (GETT) (25); and the Frueh and Quin framework that aims to facilitate communication between test developers and health-technology evaluators (26).

Table 1 lists the seven evaluation steps along with some comments on challenges that arise on each step. With BGD, the functionality of this framework is unclear. For example, a rapid triage step makes sense given the volume of information, but who, when and how would do this rapid triage for zillions of emerging big data and related tests is unclear. The ability to rapidly triage personalized information is unknown, but preliminary evidence suggests that the process can be arduous and even well-trained experts may reach different conclusions (9), for example about the degree of pathogenicity of specific genetic variants (27). Even with relatively simple genetic tests, guidelines of different organizations and professional societies are often disagreeing with each other (28). Repositories of guidelines and other BGD-related decisions need to be user-friendly, easy to navigate, unambiguous and objective. There is little evidence that any of these features are easy to achieve. Any effort may have a better chance of success, if instead of waiting for the retrospective accumulation of fragmented published data and scattered publically available resources, there is large-scale international collaboration with large-scale sharing of evidence and continuous updating as more data accrue.

Of the seven steps, the first may be the most important: identifying the clinical problem to solve and whether any good alternatives already exist. Past research has typically not followed a problem-based approach. Innumerable papers have accumulated in many fields without a clear rationale of why this research was done and what the aim was in terms of clinical translation. This makes the retrieval and assessment of relevant alternatives difficult. In some clinical applications, the number of existing alternatives is already stunningly large. For example, there are ongoing efforts to improve prediction with genetic information for cardiovascular disease. However, a systematic review found that 363 different predictive models already exist for cardiovascular disease (29). None seems to confer a clear advantage. As another example, the availability of big data through EHRs allows building predictive models with a much larger number of predictors than previous traditional predictive models. However, a systematic review of such models shows (30) that their discriminating ability remains modest. Starting from the clinical question that needs to be answered and getting clinicians more involved is essential to make meaningful progress (31). Simply accruing more data and using more automation may even solidify the various forms of automation bias (32) without offering a clinical advantage.

Do we need RCTs to assess the clinical utility of big genomic data?

RCTs are a centerpiece of EBM and they will continue to be important also in the BGD era. RCTs are not always easy to conduct. Therefore, different types of evidence are also useful to consider. Different types of designs may help inform about different clinical questions. However, clinical utility would be difficult to establish with high certainty in the absence of comparative evidence, and in particular RCTs.

A recent umbrella review evaluated 21 systematic reviews published between 2010 and 2015 on clinical applications of genomics (33). The authors found very limited evidence about the effect of using genomic tests on health outcomes. The systematic reviews found substantial risk of bias, and limited randomized evidence. However, this picture applies mostly to technologies and tests that predate the current massive availability of BGD. The questions need to be reassessed for new technologies.

One major question is whether genomic information can change behavior. A systematic review published in 2016 (34) identified 18 randomized and quasi randomized studies that reported on behavioral outcomes, including smoking cessation (six studies; $n = 2663$), diet (seven studies; $n = 1784$), physical activity (six studies; $n = 1704$) and less data on alcohol use, medication use, sun protection and attendance at screening or behavioral support programs. There were no apparent benefits of communicating genetic-based risk estimates on smoking cessation (odds ratio 0.92, 95% CI 0.63–1.35), diet (standardized mean difference 0.12, 95% CI, −0.00–0.24), physical activity (standardized mean difference 0.03, 95% CI −0.13–0.08), or anyother behaviors or on motivation to change behavior, and no adverse effects, such as depression and anxiety. The evidence was typically of low quality and studies were at high or unknown risk of bias. However, most studies used very limited genetic information from the candidate gene era. Table 2 summarizes RCTs that have assessed newer multigenic predictive score information [i.e. scores made of multiple (19) common genetic variants] for improving outcomes in various diseases (35–38). The results are rather disappointing. Of course, one may retort that even these genetic risk scores still used genetic signatures that carried limited information.

Of note, even outside genetics there are no more than a few hundreds of RCTs of diagnostic, prognostic, predictive and monitoring tests. A systematic review of 140 such RCTs with 153 comparisons found significant effectiveness for patient outcomes of only 28 (18%) of them, a small minority (39). Similarly, the literature on traditional screening methods shows that very few of them have well established roles with strong evidence to support them (40). Expectations from big genomics to reverse this picture of ineffectiveness for much of current diagnostic and predictive medicine should be cautious. The detrimental consequences of testing individuals without clear clinical utility may include increased cost, overdiagnosis, further diagnostic and therapeutic waste and eventually worse hard clinical outcomes.

Of course, traditional RCTs do have their own well-known biases, the discussion of which is beyond the scope of this paper (reviewed in 41,42). One can avoid most of these biases pre-emptively with careful design, conduct, analysis and reporting. Moreover, there is a new range of trial designs [including Bucket, Basket, Umbrella, Adaptive, and Sequential, Multiple Assignment, Randomized Trial (SMART) designs] (reviewed in 43) that can incorporate creatively precision information, for example tumor genomic profiles in oncology. One may also hybridize design features, for example the NCI-MATCH trial discussed above (16), is a Super Umbrella trial, combining features of Bucket and Umbrella trials. As of the writing of this paper, all these new trial designs cumulatively represent less than 1% of all ongoing oncology RCTs, and are even more uncommon in other fields, but broader use should be encouraged.

How do we design an evidence-based research agenda to maximize utility for big genomic data?

Most clinical research done to-date has not been useful (44), because clinical utility was never a top consideration. This also applies to BGD and may also apply to future efforts unless clinical utility becomes a key objective. One of us has previously proposed eight major features of useful clinical research (44). First, a solid problem must exist that needs to be solved, rather than create a problem that does not exist and add confusion with collecting irrelevant, massive BGD information. Second, context placement needs to ascertain what is already known on the question of interest based on previous BGD or other relevant data. Third, future studies need to be designed aiming for maximizing information gain, regardless of what their results might be. Fourth, pragmatism must ensure that the results apply to real life, while collection and interpretation of some types of BGD may not be straightforward outside research settings. Reliable, pragmatic action may be more important than the exact process of dissecting complex information. In some cases, a pragmatic approach may even depend on a black box, for example having an artificial intelligence system make final practical recommendations, as in the case of Watson for Oncology making recommendations for breast cancer management (45). Fifth, patient centeredness should allow patients to express their real needs and influence research to address what they do care about mostly in their lives and about their health. Sixth, value for money needs to be secured, and this can be a great challenge for expensive BGD technologies. Despite a decreasing cost per unit of testing, massive testing may result in a waste of funds and the adverse consequences of unnecessary or misleading information may cost even more. Seventh, feasibility should be assessed on a case-by-case basis, as many big data initiatives may be too ambitious. Even if they can get off the ground, there may not be sufficient resources for their subsequent maintenance. Finally, transparency should document that BGD and other data are shared and possible to verify, re-use and integrate and that they are as accurate and unbiased as possible (46). Genomics has a strong tradition at the forefront of data sharing. Nevertheless, as BGD become universal, issues of privacy also need to be handled carefully and new regulatory requirements may arise (47). Overall, while BGD hold a lot of promise for transforming medicine, we will need to generate evidence that this transformation can be evidence based. The combination of EBM and BGD may allow reaping maximum benefits.

Acknowledgments

References

1. Collins FS and Varmus H (2015) A new initiative on precision medicine. N. Engl. J. Med, 372, 793–795. [PubMed: 25635347]

2. Khoury MJ and Ioannidis JP (2014) Big data meets public health. Science, 346, 1054–1055. [PubMed: 25430753]

3. Roberts JS, Gornick MC, Carere DA, Uhlmann WR, Ruffin MT and Green RC (2017) Direct-to-consumer genetic testing: user motivations, decision making, and perceived utility of results. Public Health Genomics, 20, 36–45. [PubMed: 28068660]

4. Eddy DM (1990) Practice policies: guidelines for methods. JAMA, 263, 1839–1841. [PubMed: 2313855]

5. Sackett DL, Rosenberg WM, Gray JA, Haynes RB and Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ, 312, 71–72. [PubMed: 8555924]

6. Khoury MJ (2017) No shortcuts on the long road to evidence-based genomic medicine. JAMA, 318, 27–28. [PubMed: 28570728]

7. Ioannidis JPA, Stuart ME, Brownlee S and Strite SA (2017) How to survive the medical misinformation mess. Eur. J. Clin. Invest, 47, 795–802. [PubMed: 28881000]

8. Roberts MC, Kennedy AE, Chambers DA and Khoury MJ (2017) The current state of implementation science in genomic medicine: opportunities for improvement. Genet. Med, 19, 858–863. [PubMed: 28079898]

9. Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, Merker JD, Goldfeder RL, Enns GM, David SP et al. (2014) Clinical interpretation and implications of whole-genome sequencing. JAMA, 311, 1035–1045. [PubMed: 24618965]

10. Tonelli MR and Shirts BH (2017) Knowledge for precision medicine: mechanistic reasoning and methodological pluralism. JAMA, 318, 1649–1650. [PubMed: 29052713]

11. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D and Keller J (1988) A clinician's guide for conducting randomized trials in individual patients. CMAJ, 139, 497–503. [PubMed: 3409138]

12. Larson EB, Ellsworth AJ and Oas J (1993) Randomized clinical trials in single patients during a 2-year period. JAMA, 270, 2708–2712. [PubMed: 8133588]

13. Duan N, Kravitz RL and Schmid CH (2013) Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. J. Clin. Epidemiol, 66, S21–S28. [PubMed: 23849149]

14. Scuffham PA, Nikles J, Mitchell GK, Yelland MJ, Vine N, Poulos CJ, Pillans PI, Bashford G, del Mar C, Schluter PJ et al. (2010) Using N-of-1 trials to improve patient management and save costs. J. Gen. Intern. Med, 25, 906–913. [PubMed: 20386995]

15. Joyner MJ, Paneth N and Ioannidis JP (2016) What happens when underperforming big ideas in research become entrenched? JAMA, 316, 1355–1356. [PubMed: 27467098]

16. Conley BA and Doroshow JH (2014) Molecular analysis for therapy choice: NCI MATCH. Semin. Oncol, 41, 297–299. [PubMed: 25023344]

17. FDA expands approved use of Kalydeco to treat additional mutations of cystic fibrosis. 2017 https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm559212.htm; date last accessed November 20, 2017.

18. FDA approves first cancer treatment for any solid tumor with a specific genetic feature. 2017 https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm560167.htm; date last accessed November 20, 2017.

19. Opentracker. Definitions of big data. https://www.opentracker.net/article/definitions-big-data/; date last accessed February 26, 2018.

20. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Board on the Health of Select Populations, Committee on the Evidence Base for Genetic Testing. An evidence framework for genetic testing Washington (DC): National Academies Press (US); 2017 3 27.

21. USPSTF Methods and Processes. https://www.uspreventiveservicestaskforce.org/Page/Name/methods-and-processes; date last accessed November 20, 2017.

22. Fryback DG and Thornbury JR (1991) The efficacy of diagnostic imaging. Med. Dec. Making, 11, 88–94.

23. Centers for Disease Control and Prevention (2013) ACCE model list of 44 targeted questions aimed at comprehensive review of genetic testing. http://www.cdc.gov/genomics/gtesting/acce/acce_proj.htm; date last accessed November 20, 2017.

24. EGAPP (2014) The EGAPP initiative: lessons learned. Genet. Med, 16, 217–224. [PubMed: 23928914]

25. Rousseau F, Lindsay C, Charland M, Labelle Y, Bergeron J, Blancquaert I, Delage R, Gilfix B, Miron M, Mitchell GA et al. (2010) Development and description of GETT: a genetic testing evidence tracking tool. Clin. Chem. Lab. Med, 48, 1397–1407. [PubMed: 20658948]

26. Frueh FW and Quinn B (2014) Molecular diagnostics clinical utility strategy: a six-part framework. Expert Rev. Mol. Diagn, 14, 777–786.

27. Manrai AK, Ioannidis JP and Kohane IS (2016) Clinical genomics: from pathogenicity claims to quantitative risk estimates. JAMA, 315, 1233–1234. [PubMed: 26925924]

28. Chang CQ, Tingle SR, Filipski KK, Khoury MJ, Lam TK, Schully SD and Ioannidis JP (2015) An overview of recommendations and translational milestones for genomic tests in cancer. Genet. Med, 17, 431–440. [PubMed: 25341115]

29. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C et al. (2016) Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ, 353, i2416. [PubMed: 27184143]

30. Goldstein BA, Navar AM, Pencina MJ and Ioannidis JP (2017) Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J. Am. Med. Inform. Assoc, 24, 198–208. [PubMed: 27189013]

31. Mandl KD and Bourgeois FT (2017) The evolution of patient diagnosis: from art to digital data-driven science. JAMA, 318, 1859–1860. [PubMed: 29075757]

32. Goddard K, Roudsari A and Wyatt JC (2012) Automation bias: a systematic review of frequency, effect mediators, and mitigators. J. Am. Med. Inform. Assoc, 19, 121–127. [PubMed: 21685142]

33. Phillips KA, Deverka PA, Sox HC, Khoury MJ, Sandy LG, Ginsburg GS, Tunis SR, Orlando LA and Douglas MP (2017) Making genomic medicine evidence-based and patient-centered: a structured review and landscape analysis of comparative effectiveness research. Genet. Med, 19, 1081–1091. [PubMed: 28406488]

34. Hollands GJ, French DP, Griffin SJ, Prevost AT, Sutton S, King S and Marteau TM (2016) The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis. BMJ, 352, i1102. [PubMed: 26979548]

35. Godino JG, van Sluijs EM, Marteau TM, Sutton S, Sharp SJ and Griffin SJ (2016) Lifestyle advice combined with personalized estimates of genetic or phenotypic risk of type 2 diabetes, and objectively measured physical activity: a randomized controlled trial. PLoS Med., 13, e1002185. [PubMed: 27898672]

36. Grant RW, O'Brien KE, Waxler JL, Vassy JL, Delahanty LM, Bissett LG, Green RC, Stember KG, Guiducci C, Park ER et al. (2013) Personalized genetic risk counseling to motivate diabetes prevention: a randomized trial. Diabetes Care, 36, 13–19. [PubMed: 22933432]

37. Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, Haddad RA, Marroush TS, Shameer K, Olson JE et al. (2016) Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). Circulation, 133, 1181–1188. [PubMed: 26915630]

38. Knowles JW, Zarafshar S, Pavlovic A, Goldstein BA, Tsai S, Li J, McConnell MV, Absher D, Ashley EA, Kiernan M, Ioannidis JPA et al. (2017) Impact of a genetic risk score for coronary artery disease on reducing cardiovascular risk: a pilot randomized controlled study. Front. Cardiovasc. Med, 4, 53. [PubMed: 28856136]

39. Siontis KC, Siontis GC, Contopoulos-Ioannidis DG and Ioannidis JP (2014) Diagnostic tests often fail to lead to changes in patient outcomes. J. Clin. Epidemiol, 67, 612–621. [PubMed: 24679598]

40. Saquib N, Saquib J and Ioannidis JP (2015) Does screening for disease save lives in asymptomatic adults? Systematic review of meta-analyses and randomized trials. Int. J. Epidemiol, 44, 264–277. [PubMed: 25596211]

41. Ioannidis JP (2014) Clinical trials: what a waste. BMJ, 349, g7089. [PubMed: 25499097]

42. Gluud LL (2006) Bias in clinical intervention research. Am. J. Epidemiol, 163, 493–501. [PubMed: 16443796]

43. Biankin AV, Piantadosi S and Hollingsworth SJ (2015) Patient-centric trials for therapeutic development in precision oncology. Nature, 526, 361–370. [PubMed: 26469047]

44. Ioannidis JP (2016) Why most clinical research is not useful. PLoS Med, 13, e1002049. [PubMed: 27328301]

45. Somashekhar SP, Sepúlveda MJ, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C, Rauthan A, Arun Kumar N, Patil P, Rhee K et al. (2018) Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. Ann. Oncol, 45, 418–423.

46. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers E-J, Ware JJ and Ioannidis JP (2017) A manifesto for reproducible research. Nat. Human Behav, 1, 0021.

47. Lipworth W, Mason PH, Kerridge I and Ioannidis JP (2017) Ethics and epistemology in big data research. J. Bioeth. Inq, 14, 489–500. [PubMed: 28321561]

Table 1.

An evidence framework for genetic testing: seven steps and some challenges for each step with the advent of BGD

Step	Challenges
Define genetic test scenarios on the basis of the clinical setting, the purpose of the test, the population, the outcomes of interest and comparable alternative methods	Unfortunately most research to-date has not used this approach of starting from the clinical problem that needs to be solved, thus there is little evidence on comparable alternative methods that fits to this framework
For each genetic test scenario, conduct an initial structured assessment to determine whether the test should be covered, denied, or subject to additional evaluation	It is unclear what exactly this initial structured assessment would entail, if it not a full systematic review. While method for rapid reviews and scoping reviews do exist or get developed, it is unclear how well they would work in the case of big genomics. The proposed step seems like an effort to quickly get rid of a large number of tests in an environment where there would be a difficult to handle mass of big data, but it is unclear if cutting corners will help or make things worse
Conduct or support evidence-based systematic reviews for genetic test scenarios that require additional evaluation	Welcome emphasis on systematic review approach. Systematic reviews however, have major problems when conducted retrospectively with fragmented data subject to publication biases. Given the strong tradition of genetics in data sharing, there is an opportunity to promote a model of large-scale international collaboration with prospective, ongoing, continuously updated reviews, as new data accumulate
Conduct or support a structured process to produce clinical guidance for a genetic test scenario	This clause anticipates that there are many contextual issues that go beyond the strict evidence review, for example social issues, net benefits and harms, and aggregate costs. Some of these may be difficult to define, they may be setting-dependent, and they may carry substantial subjectivity. There is extensive evidence about how to produce guidelines and also about caveats in the process. Given the massive information, generating and updating guidelines for BGD will be a major challenge
Publicly share resulting decisions and justification about evaluated genetic test scenarios, and retain decisions in a repository	A repository is useful to the extent that it can be comprehensive, systematic and also allow user-friendly navigation so that one can readily find the most appropriate guidance. Experience with traditional guidelines repositories exist (e.g. the Guidelines Clearinghouse), but it is unclear if the same concept would work with the massive and rapidly evolving BGD
Implement timely review and revision of decisions on the basis of new data	As above, this would have the best chances of success, if evidence is incorporated in real-time based on some international collaboration and sharing with accumulation of all relevant data. Still, reviewing and revising all decisions will require enormous resources and it is questionable whether the process can be automated and objective or will continue to require subjective calls
Identify evidence gaps to be addressed by research	This is a traditional major role of systematic reviews. More reliable and up-to-date systematic reviews would have the best chance to do this task well

Table 2.

Randomized controlled trials on providing multigenetic score information versus control without such genetic score information

Author (ref)	N	SNPs	Phenotype	Main results
Godino (35)	580	23	Type 2 DM	No effect on physical activity, self-reported diet, self-reported weight, worry and anxiety
Grant (36)	116	36	Type 2 DM	No effect on self-reported motivation, program attendance or mean weight loss
Kullo (37)	216	28	CHD	Improved LDL-C, more statin uptake, no effects on diet or physical activity
Knowles (38)	94	19	CHD	No effects on LDL-C, other CHD risk factors, weight loss, diet, physical activity, risk perceptions, and psychological outcomes