Chapter 1
Wellness and Health Omics Linked to the Environment: The WHOLE Approach to Personalized Medicine

Greg Gibson

Abstract The WHOLE approach to personalized medicine represents an effort to integrate clinical and genomic profiling jointly into preventative health care and the promotion of wellness. Our premise is that genotypes alone are insufficient to predict health outcomes, since they fail to account for individualized responses to the environment and life history. Instead, integrative genomic approaches incorporating whole genome sequences and transcriptome and epigenome profiles, all combined with extensive clinical data obtained at annual health evaluations, have the potential to provide more informative wellness classification. As with traditional medicine where the physician interprets subclinical signs in light of the person's health history, truly personalized medicine will be founded on algorithms that extract relevant information from genomes but will also require interpretation in light of the triggers, behaviors, and environment that are unique to each person. This chapter discusses some of the major obstacles to implementation, from development of risk scores through integration of diverse omic data types to presentation of results in a format that fosters development of personal health action plans.

It is a truth universally acknowledged that personal genome sequences will be a core component of individualized health care in the coming decades [23]. It is rightly claimed that genomic medicine should be predictive, personalized, preventive, and participatory, meaning that individuals will be encouraged to understand their own health risks and take preemptive measures to avert the onset of disease [5, 25]. Yet expert geneticists are debating whether genotypic predictors are now or will ever be more predictive than family history and clinical indicators [17, 28, 62], and there is reasonable skepticism surrounding causal inference from rare deleterious variants

G. Gibson (✉)
Georgia Institute of Technology, Predictive Health Institute and School of Biology,
Atlanta, GA 30332, USA
e-mail: greg.gibson@biology.gatech.edu

[37, 39]. So while tremendous progress is being made toward routine incorporation of whole genome sequence analysis for rare congenital disorders detected at birth [4, 22, 40, 56], and in cancer diagnosis and prognostics [9, 10, 18, 51], broader application to the complex common diseases that eventually afflict most adults remains to be introduced. The purpose of this chapter is to argue that the gap between hype and reality [53] needs to be addressed on two fronts: recalibration of expectation from prediction to classification and incorporation of functional genomic data into integrative predictive health.

The WHOLE approach encoded in the acronym for Wellness and Health Omics Linked to the Environment also places emphasis on the concept of wellness. Whereas the focus of most western medicine is on curing illness, universal public health strategies should attend more to disease prevention. As one of the leaders of this movement, Dr. Ken Brigham at Emory University remarks [7], the goal of health care should be to assure that "as many of us as possible should age with grace and die with painless dignity of natural causes." Our vision at the Center for Health Discovery and Well Being in Atlanta [6, 47] is that genomic data will be integrated into primary medical care precisely for this purpose, to help people make better lifestyle choices that promote the maintenance of good health.

There are three major challenges we see that need to be confronted, which are discussed successively below. The first is the development of genomic classifiers that explain a sufficient proportion of the variance in disease risk to be informative. In the near future, these will be genotype based, incorporating rare and common variants, clearly utilizing advanced statistical methodologies but also requiring adjustment for population genetic differences [15] and family structure [49]. The second is integration of sequence data with other genomic data types [19, 25, 33], such as transcriptomic, epigenomic, and metabolomic profiles, as well as with relevant clinical and biochemical measures and family history data. Whether or not the environment can be directly incorporated as well is an open question [2], though it can be argued that functional genomic data captures lifetime environmental exposure indirectly. The third challenge is working out how to present all of this data to healthy adults in a manner that is understandable and sufficiently actionable that they will commit to positive health behaviors. To this end I conclude with an outline of one strategy that is likely to involve the training of a new generation of professional genomic counselors.

1.1 Genomic Classifiers

The foundation of genomic classification is always likely to be genotypic. Single nucleotide polymorphisms identified through genome-wide association studies [28] or classical candidate gene approaches can be combined to more accurately discriminate cases and controls than single locus classifiers [61, 63]. The simplest multivariate scores are allelic sums, where the number of alleles that is associated with disease is tallied across all identified loci. For n loci, the score theoretically ranges from zero to $2n$, and the distribution is normal, but it will be skewed as a

function of the allele frequency spectrum. Few individuals will have extreme values, but under a liability threshold model, it is assumed that individuals with scores at the top of the range are at the most elevated risk of disease.

Risk can be modeled as a function of the score as a predictor in the same sense as Framingham risk scores predict likelihood of onset of disease in a given time period [8, 59], or more simply individuals above and below an appropriate value can be classified as high or low risk. For type 2 diabetes, a classifier based on 18 loci established that individuals with the top 1 % of simple allelic sum scores (25 or more risk alleles) have quadruple the risk relative to the bottom 2 % (fewer than 12 risk alleles) and slightly more than double the risk of the general population [30; see also 58]. This measure only marginally improves on the Framingham risk score for diabetes [60] and alone does not approach it for predictive power. However, at least in our CHDWB study the two measures (allelic sum and FRS) are only mildly correlated (unpublished observation), and so it is interesting to ask whether extreme genotype scores may suggest an alternative mode of diabetes risk.

A slightly more sophisticated approach is to weight the allelic scores by the magnitude of their effect. If one allele has a relative risk of 1.4, then it should have twice the impact of one with a relative risk of 1.2. In practice, it is not clear that weighted allelic sums improve on simple ones (Fig. 1.1a), perhaps reflecting the small amount of variance explained by current models built with variants that in general collectively explain no more than 20 % of disease risk. There is also likely to be large error in the estimation of individual allelic effects both due to sampling biases and incomplete LD between tagging SNPs and unknown causal variants. Nevertheless, for type 1 diabetes, a multiplicative allelic model based on 34 loci that collectively explain 60 % of the expected genetic contribution has been introduced [14, 44]. A score with a sensitivity of 80 % is achieved in 18 % of the population even though only less than half of one percent is type 1 diabetic. However, the positive predictive value remains fairly low since the false positive rate still exceeds 90 %. It seems that for rare diseases (less than 1 % of the population), it is unlikely that genotypic measures will ever be predictive in a clinical setting. Nevertheless, as a screening tool, there may be enormous financial and medical value in focusing resources on the highest risk portion of the population and excluding those least at risk from unnecessary surveillance or treatment.

If allele sums are used, it also makes sense to attempt to weight scores by allele frequencies. Two individuals may have the same score, but if one of them has most of the risk attributed to alleles that are not typically the risk allele in the population, whereas the other has the common high-risk variants, then it stands to reason that the former is likely to be at elevated overall relative risk. This is illustrated in Fig. 1.1b. An obvious way to achieve the weighting is to convert relative risks into odds ratios, compute the log sum of those odds, and regenerate a probability of disease [35]. Starting with a baseline risk for the relevant gender, ethnicity, and age group, each successive allele adds to or subtracts from the log odds, which are a function of the allelic effect and frequency.

The immediate problem with this approach is that it is susceptible to variation in allele frequencies among populations. Two people with identical weighted allelic sums may nevertheless have very different relative risks according to whether they are, for example, of African, Asian, or European descent (Fig. 1.1c). Somewhat
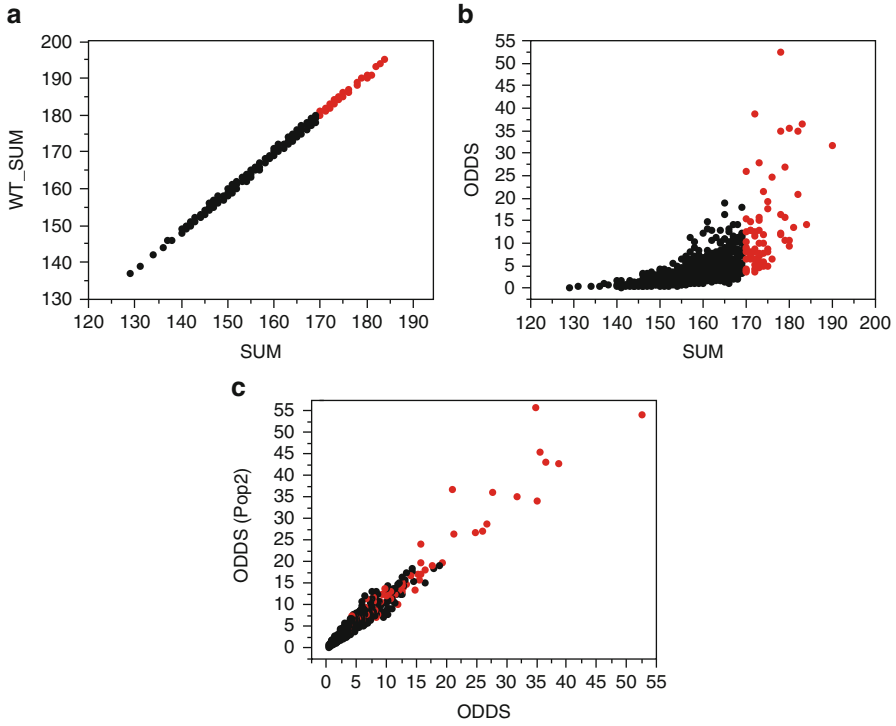
Fig. 1.1 Comparison of risk scores. The three *x*–*y* plots compare risk scores generated by three different methods, applied to a simulated dataset consisting of 200 disease SNPs measured in 1,000 people. The alleles range in risk allele frequency from 0.1 to 0.9 with a bias toward lower frequencies, and effect sizes were drawn from a normal distribution with mean of zero and standard deviation of 0.07. (**a**) Comparison of simple allelic sum score and weighted allelic sum score, showing a modest effect of weighting the sum by the effect size. Red points highlight individuals in the top decile of scores. (**b**) Comparison of simple allelic sum score and probability calculation from odds ratios obtained following the method in Morgan, Chen, and Butte [35] which computes the probability of disease from the summation of log odds ratios that are necessarily conditioned on the allele frequency. Despite increased variance of the score reflecting the multiplicative nature of the risk assessment (due to summation of log odds), the correlation in ranks is strong. (**c**) Comparison of probability scores for the same data as in (**b**) with computations assessed after randomizing the frequencies of one-quarter of the alleles, showing how population structure potentially affects disease risk assessment even where allelic effect sizes are assumed to be constant

paradoxically, heterozygosity at a single contributing locus can either increase or decrease the odds in different ethnicities, according to whether the risk allele is rare or common in either population. Accommodations can be made by deriving separate multi-allelic scores for each ethnicity, but an additional complication arises where admixture (population mixing) exists, which is the norm in contemporary America at least. Perhaps risk scores should be adjusted by the allelic frequencies expected of individuals with the observed mixture of ethnicities, but a case for local ancestry adjustment with phased genomes can be made [54, 55], and then the issue of the

appropriate baseline prevalence arises. It is not yet clear how much of an issue this is, and clearly much more research needs to be done, likely also including attention to geographically structured cultural and environmental modifiers of prevalence.

Finally, predictors and classifiers that do not assume additive effects of GWAS hits are being introduced. Sparse factorization and machine-learning approaches offer very powerful approaches that generate scores, incorporating SNPs that do not have strong univariate associations, or whose effects are conditional on other terms in the model [1, 3, 24, 29]. Often scores are developed purely as mathematical abstractions, though the interpretation is that they incorporate cryptic epistasis (genotype-by-genotype interactions) as well as environment or gender-specific interactions [48]. In these cases, there is always the assumption that the conditions and effects are consistent across populations. Again, it is not yet clear how reliable this assumption is and hence how transitive machine-learning based scores typically will be.

1.2   Integrating Functional Genomic and Clinical Data to Capture Environmental Contributions

Irrespective of the nature of the risk score, the second major challenge is to combine these into an overall personal health profile. A key insight is that the extensive comorbidity of diseases establishes the expectation that genotypic risks should covary [42, 52]. Given risk scores for dozens or even hundreds of diseases, further mathematical manipulations may facilitate gains in prediction or classification accuracy that borrow power from across diseases. At the current stage of development of personal genomic medicine, there is insufficient data to discern robust patterns of covariance, with the exception of autoimmune diseases that share common polymorphisms [32, 46]. So long as individual disease risk scores only capture a minor fraction of the genotypic risk, they are unlikely to capture to true architecture of comorbidity, but presumably this will change as more comprehensive predictors are developed.

In the mean time, Ashley et al. [2] presented a mode of visualization of combined risk that suggests how path analyses might integrate univariate risk scores. This is reproduced in Fig. 1.2b focusing just on a half dozen common disease conditions mostly related to metabolic syndrome. On the left (Fig. 1.2a), the so-called risk-o-grams [2, 13, 16] show how baseline risk for these conditions is modified by a hypothetical individual's genotypic risk. The point estimates should not be over-interpreted, the more important information being contained in the sign and magnitude of the genetic contribution. These are modified by comorbidity and redrawn in the form of the size of the font on the right, where larger circles represent increasingly elevated risk due to the individual's genotypes and the disease interactions. Interrelated disease conditions are connected by directed edges where, for example, the likelihood of developing cardiovascular disease is increased by the person's elevated risk of obesity but decreased by their low hypertension risk. Unfortunately, we do not yet have the tools to estimate the strengths of the connections, and much theoretical work on the optimal multivariate integration strategy remains to be performed.
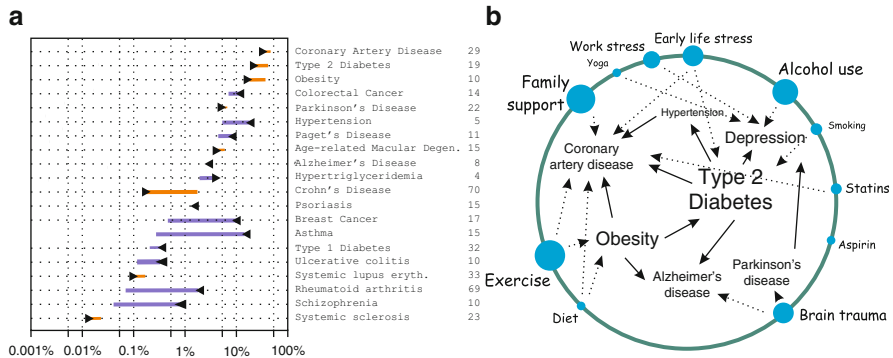
Fig. 1.2 Risk-o-grams. Following Ashley et al. [2], a hypothetical risk-o-gram (**a**) shows how genotypic risk can be used to generate a point estimate of probability of disease conditioned on the population prevalence. The figure shows a hypothetical risk assessment on the log scale for 20 diseases where the black triangles show the prevalence for the individual's gender, ethnicity, and age group, pointing to the right if genotype is predicted to increase risk or left if it decreases risk relative to the population average. The horizontal bars show the degree of genotypic effect, where, for example, Crohn's disease risk is highly elevated, but asthma and breast cancer are reduced. (**b**) These risks need to be combined, recognizing the comorbidity matrix of disease and the influence of environmental factors, including dietary and psychological stressors, exercise patterns and drug usage, and personal history of illness. The modified risk for each condition conditioned on the matrix of influences is represented by the size of the font. Although we are a long way from being able to generate robust assessments, the figure implies that classification into high- and low-risk classes should be feasible in the near future

Just as importantly, the grand circle surrounding the disease prediction network shows that the environment must also be incorporated into computations. In this case, the individual's heavy alcohol usage and lack of exercise also increase their risk of metabolic syndrome, as does a history of early life stress coupled with low family support and high work pressure. It is apparent that they are already taking statins and eating a low-fat diet to offset some of the risk, and regular yoga practice may help qualitatively. A traumatic brain injury suffered in a car accident as a child may have been a trigger that cannot be factored into population-based measures of risk, but it also feeds into likely cognitive decline with age. Again, it is not yet obvious how these environmental risks should be formulated from a statistical perspective. Drug usage can conceivably be incorporated as a cofactor in the computation of individual risk scores, but it is less obvious how to model diet and mental stress, or what the appropriate multivariate framework may be. A further advantage of this visualization is that it readily lends itself to dynamic representation of how lifestyle modifications may reduce the risk of key diseases, as individuals can observe projected changes in risks if they adopt new health behaviors.

Another aspect of the environment that we may endeavor to incorporate is cultural and geographic differentiation. Perusal of the Centers for Disease Control (CDC) database of morbidity (see, e.g., http://www.cdc.gov/cancer/dcpc/data/state.htm for cancer data) shows that most diseases have very different prevalence

according to the location within the United States. An excellent example is the well-known southern stroke belt [11] stretching from Louisiana across Alabama to Georgia and the Carolinas, but cancer incidence and many other diseases vary from region to region. Undoubtedly, rural and urban lifestyles impact disease risk, and we have shown that they also impact peripheral blood gene expression profiles [26, 36], while emerging data also suggests differences in the microbiome [64]. Most readily, this type of information could be incorporated into risk prediction already at the level of baseline prevalence, which might be assessed regionally rather than simply by gender and ethnicity. Of course someone who moves from Manhattan, New York, to Manhattan, Kansas, does not modify their risk overnight, so yet another obstacle to absolute risk prediction lies in assessing the perdurance of lifestyle effects and the impact of life stage. Notably, there is accumulating evidence that early life stress is among the biggest risk factors for a wide range of diseases, particularly in lower socioeconomic strata [21, 34, 41].

Another unresolved issue is to what extent genotype-by-environment interactions need to be taken into account in risk evaluation. There is very little evidence from GWAS that G×E is either prevalent or of sufficient magnitude to be important components of population variance [57], notwithstanding occasional reports, for example, of smoking by nicotinic acetylcholine receptor polymorphism interactions with lung cancer [65] and of arsenic by solute carrier interactions for bladder cancer [27]. This is surprising given the prevalence of both genotypic and environmental effects on gene expression [26]. Supposing that low transcript abundance for a particular gene in a relevant tissue contributes to disease risk, those homozygous for a low expression *cis*-regulatory polymorphism, in an environment where expression is significantly reduced as well, will constitute the most at-risk group. Under a liability model, G×E for disease is plausible, even in the absence of interaction effects between the genotype and gene expression. However, large eQTL effects do not translate into large disease effects measured in case-control GWAS settings. It is possible that genotypic risk score-by-environment interactions will be observed, but such studies are yet to be performed. Furthermore, perhaps the more important mode of interaction is with individualized effects, such as triggers (accidents, transient stresses) that either are not captured in epidemiological surveys or have such high variance that interaction effects do not attain significance in population-scale studies.

All of these considerations add uncertainty to risk assessment and raise the question of whether it might not be better to measure the impact of the environment biochemically. The notion is that a person's individuality results from the longitudinal interaction of their genome with all of the above lifestyle and environmental factors. These influences mediate disease risk ultimately by modifying metabolism and physiology, which in turn are a function of gene expression, which is subject to epigenetic modification. Consequently, measurement of the metabolome, transcriptome, and epigenome (e.g., chromatin methylation) should provide parallel omic information of high relevance to health care [25]. This systems biology approach is much hyped [53], but many would argue that it has yet to provide the clinical or mechanistic insights that have stemmed from genotype and sequence-based genomic medicine. A major limitation of course is that only a few tissues, principally

peripheral blood or sometimes adipose biopsy, are readily available for high-throughput analysis. Blood does reflect immune and metabolic function and possibly mirrors psychological stressors [20, 31], so there is undoubtedly much to be learned from characterization of the sources of variance, and major advances in predictive health can be expected from this approach in the next decade.

Just to briefly highlight two strategies from our own work. First, characterization of extremes of individual transcript abundance detected by either microarray or RNASeq analysis of individuals is in many ways equivalent to rare deleterious coding variant detection from sequencing. We do not yet know how to read regulatory variation directly, but this is unnecessary if it can be directly demonstrated that an RNA (or protein) is not expressed in a particular individual. Association of such differential expression with phenotypes is subject to the same caveats as rare variant association analysis. Second, transcriptional variation is highly structured and characterized by major axes that represent aspects of lymphocyte function such as B and T cell signaling, antiviral responsiveness, and inflammation [45]. This variability is evident in the principal components of peripheral blood gene expression, but also appears in modules and axes of variation that are captured by the expression of biomarker genes [12], or blood informative transcripts. We postulate that the level of activity of gene expression in these axes will be found to correlate with aspects of immune and metabolic health.

1.3   Presenting and Interpreting Genomic Risk for Wellness

The third great challenge is to present genomic indicators of disease risk to healthy individuals in a manner that will help them to make sensible health behavior choices. This is one of the major goals of the emerging discipline of medical informatics. Risk-o-grams (Fig. 1.2a) are an excellent starting point since they present risk both in absolute terms as well as apportioning the genetic contribution relative to the population average. However, they have some obvious drawbacks, not least of which is the overwhelming number of assessments, many of which are for rare conditions or are clinically not actionable. They also fail to convey a sense of the error associated with risk assessments: we are used to the notion that heavy smoking more than doubles your lifetime risk of lung cancer, yet know heavy smokers who never get lung disease and never-smokers who do. Inevitably inappropriate presentation of genetic risks will engender skepticism toward genomic medicine that may undermine the certain benefits that stand to be realized.

For this reason, in the context of wellness, classification is the more appropriate emphasis than prediction. Classification into very-high-, high-, normal-, low-, and very-low-risk levels should help individuals to focus on those aspects of their health that will benefit from close attention. It draws attention away from the myriad statistical issues discussed above, instead promoting joint consideration of genetic and clinical measures. Furthermore, it is consistent with a simplification of risk presentation in health domains that recognize patterns of comorbidity and leverage existing
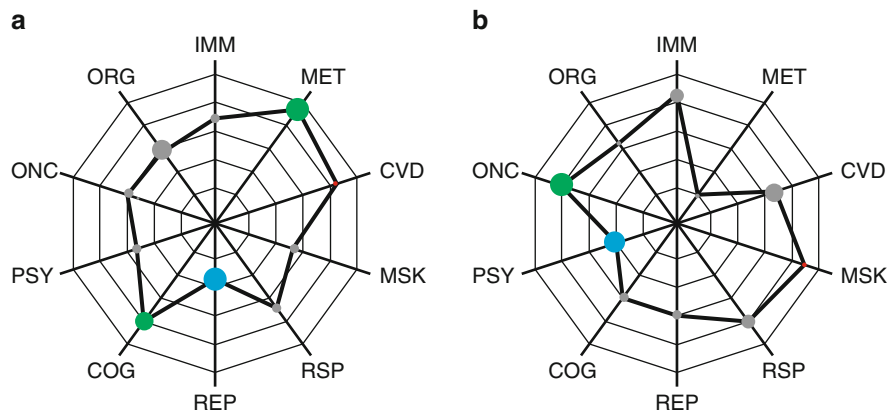
Fig. 1.3 Spider-web plots representing genomic and clinical risk in ten health domains for two hypothetical individuals. Genomic risk scores, generated by combining genotypic and functional genomic evidence, place each person in one of five risk classifications from very high (*outer band*) to very low (*inner band*) in ten health domains (*IMM* immunological, *MET* metabolic, *CVD* cardiovascular, *MSK* musculoskeletal, *RSP* respiratory, *REP* reproductive, *COG* cognitive, *PSY* psychiatric, *ONC* oncological, *ORG* organ failure). Clinical risk assessments generated from comprehensive medical examinations as well as personal and family history of disease are indicated by the size of the dots in each axis. Colors represent discordance between genomic and clinical risk as these situations are likely to be of greatest interest for individuals, alongside concordance for high risk, as they develop health action plans. Details and actual individual examples are described in Patel et al. [43]

modes of health assessment. At the Center for Health Discovery and Well Being, we are promoting the idea that comprehensive clinical evaluation annually, starting in the fourth decade of life, will foster prevention over reaction as it engages individuals in their own health choices [43]. Figure 1.3 suggests one mode of presentation of genomic data that may be incorporated into the preventative medicine framework.

Each radiating axis on the spider-web plots represents one of ten health domains. The bold polygon crosses each axis at a point, representing genomic risk in that domain (points further out mean higher risk), while the size of the circle at that point represents the observed clinical risk and/or evidence for disease. A quick glance at the spider-web plot tells an individual where they have high or low genetic and clinical risk. Areas of continuity between genetics and clinical data are highlighted as green dots. Discontinuities may be even more interesting. Those indicated in red where genetic risk is high but there is no sign of clinical danger (cardiovascular disease for A and musculoskeletal decay for B) suggest situations where the individual may pay close attention despite current good health. By contrast, situations where the genetic risk is low but clinical signs are not hopeful (respiratory disease for the smoker A and psychiatric problems for the socially isolated person B) may suggest that lifestyle changes are likely to have an impact. The main objective of this combined genomic and clinical classification is not to predict disease but to help individuals focus attention on areas where they should concentrate their health-related behaviors and surveillance.

The proposed ten common health domains are as follows:

- Immunological, including autoimmune (type 1 diabetes, multiple sclerosis, SLE, arthritis), inflammatory (especially bowel diseases), and infectious (viral and microbial) disease susceptibility, many of which show comorbidity and all of which should be related to gene expression in various blood cells
- Metabolic syndrome, generally referring to obesity and either hyperlipidemia or high blood glucose, leading to type 2 diabetes, and encompassing impaired insulin production and sensitivity
- Cardiovascular, primarily atherosclerosis and hence related to metabolic dysfunction, but also including cardiomyopathy, arrhythmia, and heightened risk of myocardial infarction or stroke
- Respiratory discomfort, namely, asthma, COPD, and fibrosis, all of which are exacerbated by smoking and call for attention to genotype-by-environment interaction
- Musculoskeletal problems, such as low bone density, chronic back pain, and muscle weakness or wasting, which are a primary cause of reduced quality of life for large percentage of the elderly
- Mental health, manifesting as depression and/or anxiety in an increasingly alarming percentage of adults, but also including schizophrenia, autism spectrum, and attention deficit disorders in adolescents and young adults
- Cognitive decline, whether due to Alzheimer's disease, Parkinson's disease, or generalized senile dementia and expected to become the major public health burden of the twenty-first century
- Cancer risk, assessed from family history and possibly peripheral blood biomarkers
- Organ malfunction, which is unlikely to have a common genomic foundation but collectively loss of eyesight, hearing, and renal and liver function, are a major source of morbidity
- Reproductive health, namely, the capacity to conceive and maintain pregnancy or to produce fertile sperm, but also including endometriosis and other causes of uterine discomfort

Pharmacological variation, for both toxicity and responsiveness to specific drugs, is also an important aspect of genomic health, sometimes having a simple genetic basis (e.g., warfarin [50]) but generally as complex as disease risk [38]. This is not by any means an exhaustive list of disease but is meant to capture the major domains that concern adults as they enter middle age and begin to make lifestyle modifications in response to self-perception of personal health concerns. Genome-wide association studies have been performed for specific diseases in each domain, and thousands of variants are available for generation of risk scores. Similarly, relevant clinical measures can be taken during routine medical checkups or as part of a dedicated wellness program such as the CHDWB and collectively generate risk profiles in these ten domains as well.

An immediate concern is how to collapse disparate genotypic and clinical risk scores into summary measures of risk for the various domains. For clinical measures, z-scores place each person in relative risk categories with those within one

standard deviation of the mean being at intermediate risk, those between 1 and 2 standard deviations at high (or low) risk, and everyone at the extremes at the very-high- or low-risk categories. A similar strategy could be applied to genotypic risk, or thresholds can be established based on the risk score distributions. Geometric means might be used to combine multiple scores, enhancing the relevance of individual high-risk values. My concern here is not with the optimal mode of collapsing but rather to suggest how spider-plot or similar visualization might be interpreted.

After consulting the spider-web plot with a physician or other health professional, the next step would be to examine the contributing risk factors in more detail. Consider the three examples. (1) In the cardiovascular domain, individual B in Fig. 1.3 has intermediate overall risk, but close examination shows that she is discordant for high blood pressure and lower than average genotypic risk of hypertension. This may suggest that some aspect of lifestyle, either high levels of job stress or a high salt diet, is responsible, and the low genetic risk might in some cases provide impetus for the individual to address the root cause. (2) Person A is concordant for obesity and high genetic risk of obesity, both of which produce high scores in the metabolic domain. Rather than accepting this as a fait accompli, with appropriate counseling she may learn that much of the genetic risk is due to neurological factors rather than any deficit in metabolic enzyme function, and this may help him to seek guidance in controlling dietary compulsions. (3) Another individual may be discordant in the organ failure domain for high genetic risk of age-related macular degeneration, but as a 70-year-old with above average eyesight has paid no attention to the possibility that he may soon suffer from loss of vision. Knowing the genetic risk, he will now have regular eye exams and follow emerging guidelines directed at preventing onset of the disease.

As discussed earlier, I envisage that genomic risk assessment will eventually incorporate transcriptional, epigenomic, and metabolic measures. The costs involved will be an obstacle for the foreseeable future, and it is not clear who will pay. It is nevertheless not difficult to see how a few thousand dollars spent on genomic analyses in middle age may save tens or hundreds of thousands of dollars in acute medical care for people approaching retirement age. Employers stand to benefit from reduced absenteeism and elevated productivity, and economic modeling suggests that the savings can be substantial. Scientific demonstration of the clinical efficacy of joint genomic and clinical profiling will likely take thousands of case studies over several years, a daunting challenge, but given the stakes, one that must be taken on.

1.4   Conclusion

Assuming success of the WHOLE paradigm, there will also be a need for training of a new class of health-care professional. A few genetic counseling programs are beginning to provide training in the interpretation of genome sequences. At the CHDWB, we have developed a Certificate program for Health Partners who consult with participants on the interpretation of their clinical profiles and help them to formulate personal health action plans. The combination of advanced genetic

counseling with a health partner is expected to yield genomic counselors, masters level professionals who will work alongside physicians, dieticians, personal trainers, and clinical geneticists to provide people who care to take advantage of the wealth of information implicit in genomic medicine, with a path to health maintenance and extended well-being.

References

1. Abraham G, Kowalczyk A, Zobel J, Inouye M (2012) Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. Genet Epidemiol 2012 Epub ahead of print
2. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE et al (2010) Clinical assessment incorporating a personal genome. Lancet 375:1525–1535
3. Ayers KL, Cordell HJ (2010) SNP selection in genome-wide and candidate gene studies via penalized logistic regression. Genet Epidemiol 34:879–891
4. Bick D, Dimmock D (2011) Whole exome and whole genome sequencing. Curr Opin Pediatr 23:594–600
5. Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF et al (2011) Systems medicine and integrated care to combat chronic noncommunicable diseases. Genome Med 3:43
6. Brigham KL (2010) Predictive health: the imminent revolution in health care. J Am Geriatr Soc 58(Suppl 2):S298–S302
7. Brigham KL, Johns MME (2012) Predictive health: how we can reinvent medicine to extend our best years? Basic Books, New York, NY
8. Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, Ebrahim S (2003) Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. BMJ 327:1267
9. Cancer Genome Atlas Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474:609–615
10. Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330–337
11. Casper ML, Wing S, Anda RF, Knowles M, Pollard RA (1995) The shifting stroke belt: changes in the geographic pattern of stroke mortality in the United States, 1962 to 1988. Stroke 26:755–760
12. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C et al (2008) A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 29:150–164
13. Chen R, Butte AJ (2011) The reference human genome demonstrates high risk of type 1 diabetes and other disorders. Pac Symp Biocomput 2011:231–242
14. Clayton DG (2009) Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 5:e1000540
15. de Roos AP, Hayes BJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. Genetics 183:1545–1553
16. Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C et al (2011) Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet 7:e1002280
17. Do CB, Hinds DA, Francke U, Eriksson N (2012) Comparison of family history and SNPs for predicting risk of complex disease. PLoS Genet 8:e1002973
18. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ et al (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486:353–360
19. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F et al (2008) Genetics of gene expression and its effect on disease. Nature 452:423–428
20. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M et al (2012) Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry 51:934–944

21. Gluckman PD, Hanson MA, Cooper C, Thornburg KL (2008) Effect of *in utero* and early-life conditions on adult health and disease. N Engl J Med 359:61–73
22. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ (2012) Next generation sequencing in clinical medicine: challenges and lessons for pathology and biomedical informatics. J Pathol Inform 3:40
23. Hamburg MA, Collins FS (2010) The path to personalized medicine. N Engl J Med 363:301–304
24. Han F, Pan W (2010) Powerful multi-marker association tests: unifying genomic distance-based regression and logistic regression. Genet Epidemiol 34:680–688
25. Hood L, Balling R, Auffray C (2012) Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J 7:992–1001
26. Idaghdour Y, Czika W, Shianna KV, Lee SH, Visscher PM et al (2010) Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet 42:62–67
27. Karagas MR, Andrew AS, Nelson HH, Li Z, Punshon T et al (2012) SLC39A2 and FSIP1 polymorphisms as potential modifiers of arsenic-related bladder cancer. Hum Genet 131:453–461
28. Kooperberg C, LeBlanc M, Obenchain V (2010) Risk prediction using genome-wide association studies. Genet Epidemiol 34:643–652
29. Kruppa J, Ziegler A, König IR (2012) Risk estimation and risk prediction using machine-learning methods. Hum Genet 131:1639–1654
30. Lango H; UK Type 2 Diabetes Genetics Consortium, Palmer CN, Morris AD, Zeggini E et al (2008) Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 57: 3129–3135
31. Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD et al (2009) Identifying blood biomarkers for mood disorders using convergent functional genomics. Mol Psychiatry 14:156–174
32. Lettre G, Rioux JD (2008) Autoimmune diseases: insights from genome-wide association studies. Hum Mol Genet 17(R2):R116–R121
33. Lusis AJ, Attie AD, Reue K (2008) Metabolic syndrome: from epidemiology to systems biology. Nat Rev Genet 9:819–830
34. Miller GE, Chen E, Fok AK, Walker H, Lim A, Nicholls EF, Cole S, Kobor MS (2009) Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. Proc Natl Acad Sci U S A 106:14716–14721
35. Morgan AA, Chen R, Butte AJ (2010) Likelihood ratios for genome medicine. Genome Med 2:30
36. Nath AP, Arafat D, Gibson G (2012) Using blood informative transcripts in geographical genomics: impact of lifestyle on gene expression in Fijians. Front Genet 3:243
37. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE et al (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245
38. Nebert DW, Zhang G, Vesell ES (2008) From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. Drug Metab Rev 40:187–224
39. Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D et al (2012) Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. Am J Hum Genet 91:303–312
40. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB (2012) Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet 49:353–361
41. Pace TWW, Mletzko TC, Alagbe O, Musselman DL, Nemeroff CB, Miller AH, Heim CM (2006) Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress. Am J Psychiatry 163:1630–1633
42. Park S, Yang JS, Kim J, Shin YE, Hwang J, Park J, Jang SK, Kim S (2012) Evolutionary history of human disease genes reveals phenotypic connections and comorbidity among genetic diseases. Sci Rep 2:757
43. Patel CJ, Sivadas A, Tabassum R, Thanawadee P, Zhao J, Arafat D, Chen R, Morgan AA, Martin G, Brigham KL, Butte AJ, Gibson G (2013) Whole genome sequencing in support of wellness and health maintenance. Genome Med 5:58

44. Polychronakos C, Li Q (2011) Understanding type 1 diabetes through genetics: advances and prospects. Nat Rev Genet 12:781–792
45. Preininger M, Arafat D, Kim J, Nath A, Idaghdour Y, Brigham KL, Gibson G (2013) Blood-informative transcripts define nine common axes of peripheral blood gene expression. PLoS Genet 9:e1003362
46. Ramos PS, Criswell LA, Moser KL, Comeau ME, Williams AH et al (2011) A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap. PLoS Genet 7:e1002406
47. Rask KJ, Brigham KL, Johns MME (2011) Integrating comparative effectiveness research programs into predictive health: a unique role for academic health centers. Acad Med 86:718–723
48. Ritchie ND, Hahn LW, Moore JH (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150–157
49. Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE (2012) The predictive capacity of personal genome sequencing. Sci Transl Med 4:133ra58
50. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K et al (2004) Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. Nature 427:537–541
51. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM et al (2011) Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med 3:111ra121
52. Rzhetsky A, Wajngurt D, Park N, Zheng T (2007) Probing genetic overlap among complex human phenotypes. Proc Natl Acad Sci U S A 104:11694–11699
53. Salari K, Watkins H, Ashley EA (2012) Personalized medicine: hope or hype? Eur Heart J 33:1564–1570
54. Sankararaman S, Sridhar S, Kimmel G, Halperin E (2008) Estimating local ancestry in admixed populations. Am J Hum Genet 82:290–303
55. Shriner D, Adeyemo A, Rotimi CN (2011) Joint ancestry and association testing in admixed individuals. PLoS Comput Biol 7:e1002325
56. Talkowski ME, Ordulu Z, Pillalamarri V, Benson CB, Blumenthal I et al (2012) Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med 367:2226–2232
57. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90:7–24
58. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ et al (2006) Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 3:e374
59. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB (1998) Prediction of coronary heart disease using risk factor categories. Circulation 97:1837–1847
60. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB (2007) Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 167:1068–1074
61. Wray NR, Goddard ME (2010) Multi-locus models of genetic risk of disease. Genome Med 2:10
62. Wray NR, Goddard ME, Visscher PM (2008) Prediction of individual genetic risk of complex disease. Curr Opin Genet Dev 18:257–263
63. Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet 6:e1000864
64. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG et al (2012) Human gut microbiome viewed across age and geography. Nature 486:222–227
65. Yu K, Wacholder S, Wheeler W, Wang Z, Caporaso N, Landi MT, Liang F (2012) A flexible Bayesian model for studying gene-environment interaction. PLoS Genet 8:e1002482