*This is a post-peer-review, pre-copyedit version of an article published in the* Journal of Genetic Counseling*. The final authenticated version is available online at:* <http://dx.doi.org/10.1007/s10897-018-0217-9>

**“Bridge to the Literature”?**

**Third-Party Genetic Interpretation Tools and the Views of Tool Developers**

Sarah C. Nelson1, Stephanie M. Fullerton2

**Suggested running head:** Third-party genetic interpretation tools

1. Institute for Public Health Genetics, University of Washington, Seattle, Washington
2. Department of Bioethics & Humanities, University of Washington, Seattle, Washington

**Sarah Catherine Nelson, MPH\***

University of Washington

Institute for Public Health Genetics

Box 359461

Seattle, WA 98195-9461

p: 206.543.1850

f: 206.543.5881

e: sarahcn@uw.edu

\*To whom correspondence should be addressed

**Abstract**

Patients and health care consumers can obtain access to their “raw,” or uninterpreted, genetic data from direct-to-consumer genetic testing companies, researchers, or providers and pursue self-directed analysis via third-party interpretation tools. Yet relatively little is known about the nature of currently available interpretation tools or the motivations of tool developers. We conducted a structured content analysis of 23 third-party interpretation tool websites and supporting information, tracking features such as types of information returned, modes of generating and presenting that information, and privacy and security measures. We additionally conducted qualitative interviews with a subset of 10 tool developers. A majority of tools (16 of 23, or 70%) offer some type of health or wellness-related information, often extracted from publicly-available variant annotation databases. Half of those interviewed characterized their activities as “bridging” users to the scientific literature rather than interpretation, for which they gave a variety of scientific, ethical, and regulatory justifications. The scale, heterogeneity, and complexity of information available from third-party interpretation is unprecedented. While developers aim to enlighten and empower tool users, interpretation-free “bridging” to rapidly evolving databases may instead impose burdens on genetic counselors and other health care providers asked to provide further contextualization and explanation.

**Keywords**: third-party interpretation; direct-to-consumer genomic testing; variant interpretation; personal genomic testing

**Introduction**

Patients and health care consumers have unprecedented access to their “raw” or uninterpreted genetic data. Direct-to-consumer genetic testing (DTC-GT) has historically been the most common route of access, as many companies — including 23andMe, AncestryDNA, and Family Tree DNA — allow customers to download a file of their uninterpreted genotype (typically SNP array) data in addition to the more well-known interpreted reports. However, shifts in policies and norms of genomic research and medicine are creating new avenues for individuals to obtain their uninterpreted genetic data. In the clinical context, for example, the Health Insurance Portability and Accountability Act (HIPAA) direct access right established in 2014 (45 C.F.R. § 164.524) enables patients to access full laboratory records, known as the designated record set (DRS). Access to the DRS for a genomic sequencing test could include uninterpreted sequence data (Evans, Dorschner, Burke, & Jarvik, 2014; U.S. DHHS, 2016). In the research context, national and international conversations note that many participants may want and perhaps deserve access to their individual data generated in the course of research, which may including uninterpreted genetic data (Bobe, n.d.; Global Alliance for Genomics and Health, n.d.; Lunshof, Church, & Prainsack, 2014; Nelson, 2016; The Precision Medicine Initiative NIH, 2017). Therefore, as potential providers of such data, the genetics community has a professional interest in understanding the myriad ways patients and participants might seek to leverage their uninterpreted genetic information.

 One of the most likely pursuits individuals may undertake with their genetic data is self-directed interpretation and analysis via an online third-party service, a heterogeneous collection of which are currently available (Badalato, Kalokairinou, & Borry, 2017; Spector-Bagdady & Pike, 2013; Wang et al., 2017). While most existing tools were created to process raw data files from DTC-GT companies, the model of third-party interpretation, which is independent or “unbundled” from a genotyping provider/service, is not restricted to DTC-GT data. Indeed, many existing tools already accept or plan to accept file formats more common to clinical or research sequencing, such as variant call format (VCF) files. As has been observed with DTC-GT reports (Bloss, Wineinger, Darst, Schork, & Topol, 2013; Kaufman, Bollinger, Dvoskin, & Scott, 2012; van der Wouden et al., 2016), individuals are likely to bring information from third-party interpretation tools to their providers for further explication. Therefore, providers could be doubly implicated: first as enablers of raw data access and second as managers of patients seeking to self-interpret such data.

The aim of this investigation was to characterize the existing landscape of third-party interpretation tools for personal genetic data, with the broader goal of assisting genetics professionals in understanding and anticipating the outcomes of expanding raw data access and third-party tool use. To accomplish this, we conducted a structured content analysis of tool websites and supporting information for 23 tools, complemented by qualitative interviews with a subset of tool developers. Our study contributes knowledge of the operation and motivations of current third-party interpretation services, which furthers our understanding of how expanding raw data access will affect individuals, families, researchers, and providers.

**Materials and Methods**

***Dataset***

Third-party interpretation tools were identified from a range of sources including blog posts (e.g., Bettinger, 2013), web sites (“Autosomal DNA tools,” n.d., “DNA Testing Reviews,” n.d.), DTC-GT customer discussion boards, academic conferences (e.g., Erlich, Gordon, Pearson, Shee, & Pickrell, 2015), scientific literature (Greshake, Bayer, Rausch, & Reda, 2014), and from personal and professional networks. To be included, the tool had to (1) enable a user to process or analyze “raw” genetic data from one or more DTC-GT companies, (2) return some type of information to the user, and (3) be active at the time of study (July - December 2016). These criteria encompass some companies that also offer combined genotyping and interpretation services, in addition to interpretation-only products. (For example, the DTC-GT company Family Tree DNA also offers an interpretation-only “Autosomal Transfer” service.) Note criterion (2) excludes tools exclusively focused on crowdsourcing user genotype data for research, such as Open Humans (“Open Humans,” n.d.). A total of 23 tools met these criteria and were included in this study.

***Content Analysis***

Following previous systematic studies of DTC nutrigenomics (Sterling, 2008) and ancestry (Wagner, Cooper, Sterling, & Royal, 2012) tests, we conducted a structured content analysis of third-party interpretation tools identified as described in the previous section. We used a data collection form to annotate tools on a variety of features, including the “natural history” of the tool (i.e. who started the tool, when, and why); types of information available to the user; how the tool generates that information (e.g. what bioinformatics or analytical approaches and database substrates are used); data privacy and security measures; business model/funding; number and types of users; and ability for users to contribute their data (genetic and/or phenotypic) to research activities. We also tracked where tools could process other (non DTC-GT) genetic data file types, such as VCF. Data collection forms were populated primarily from information on the tool websites, supplemented with additional information such as media articles and blog posts by tool developers and users and, where available, interview data (described below).

***Interviews***

We conducted semi-structured interviews with 10 tool developers associated with 8 different tools (8/23 = 35% of tools represented). We contacted representatives from each tool up to three times, via either email, a contact form on the tool website, and/or social media (e.g., a Facebook or Twitter account associated with the tool). Developers from six tools explicitly declined to be interviewed; others either did not respond to our initial or follow-up communications. Interview questions were designed to supplement and extend the content analysis and therefore covered the same general topics. In particular, we sought deeper understanding of tool developers’ backgrounds and motivations in developing the tools. We also asked developers about their views on regulation of both interpretation-only tools and DTC-GT more broadly. Interviews were conducted remotely — via phone or Skype, recorded, and transcribed. We thematically analyzed interview transcripts using the qualitative analysis software Atlas.ti (v8.0). Codes were generated inductively from the data, including the “bridge to the literature” code discussed further below.

All research activities were reviewed and approved by the University of Washington Institutional Review Board as minimal risk human subjects research (approval #50238).

**Results**

***Content analysis***

The 23 third-party interpretation tools we studied vary significantly across multiple domains, including how user data are provided to the tool, types of information returned to users, and processes for generating and presenting that information. The tool features discussed below are summarized in Table I; additional information is provided in Online Resource 1, including tool website URL, start year, and whether the tool offers bundled genotyping plus interpretation in addition to the interpretation-only service.

*User data: input formats, retention and sharing*

Per study inclusion criteria, all tools process uninterpreted genetic data from at least one major DTC-GT company (see “Input formats” in Table I). While 14 tools require the user to have downloaded a genetic data file (*GDF*) to provide it to the tool, 9 offer 23andMe customers the option of transferring their genotypes via the *23andMe* *API* (application program interface). Five tools accept the *VCF* format common to sequencing data.

The retention and sharing of user genotype data varies between tools. For four tools, user data are *analyzed locally* (i.e., on the user’s machine or device), rather than being uploaded to a central server. The remaining tools centrally process data, but store it for varying time periods and purposes. Five tools do *not retain* user data, but rather delete it after a short, pre-defined time period (e.g., after generating and returning a report). Some tools retain data to enable the basic functions of the tool, which involve users collating data: e.g., in GEDMatch to detect and characterize familial relatedness, and in Infinome, to allow users to compare genotypic and phenotypic data with each other. Other tools retain data for research purposes: e.g., in DNA.land, which conducts academic research internally and with collaborators, and in openSNP, where user data is made publicly available under a Creative Commons (CC0) license. Several tools use aggregated data for internal research and development and *may share* aggregated data with third parties. While most tools’ privacy statements preclude sharing of individual-level data without user consent, the sharing of aggregated or de-identified user data is often done without express user permission (i.e., apart from initial agreement to the terms of service or use).

*Type of results: genetic ancestry, genealogy and relatives, health and wellness*

Similar to the offerings of DTC-GT companies, third-party tools provide three general categories of information to users: genetic ancestry (8 tools), genealogy (5 tools), and/or health and wellness (16 tools), summarized in Table I. Some tools provide information from more than one category. Below we discuss each of these three results categories in turn. Rather than attempt to give an exhaustive list of all available reports, we instead present illustrative examples from each category.

Genetic ancestry information takes a range of forms; however, the most common type of report is a breakdown of *global* (i.e., genome-wide) ancestry composition (provided by 6 tools). Interpretome offers a slightly different view of global ancestry: a principal component analysis (PCA) plot that pinpoints the user in relation to select reference populations. GEDMatch and Interpretome provide results for *local* ancestry (i.e., ancestry along chromosomal segments), often referred to as “chromosome painting.” GPS Origins generates “migration maps” intended to illustrate historical migrations of the user’s ancestral groups. Finally, some tools assign mitochondrial DNA, mtDNA (WeGene, James Lick Haplogroup Analysis) or chromosome Y, chrY (WeGene) *haplogroups*. (Note, while the Chinese company WeGene only provides ancestry reports translated into English, customers who can read Chinese can technically access the health and wellness and relative finder reports available in the full Chinese product, on the same website.)

In the genealogy and relatives results category, information is typically identification and characterization of close relatedness to other individuals who have also provided data to the tool. Three tools offer a *relative finder* feature, which can have an advantage over DTC-GT companies’ relative finder tools in that the search space need not be limited to one company’s users (e.g., a 23andMe user can find a related AncestryDNA user with GEDMatch). In addition to finding relatives, these tools also present *shared segments* both graphically and numerically (e.g., as total amount of shared centimorgans, cM). David Pike’s utilities can perform *haplotype phasing* or generate a list of *shared segments* between either a pair or trio of individuals, but only among individuals for whom the user has data (as analysis occurs locally). Relative information in Enlis Genome Personal is restricted to the ability to track the inheritance of specific variations across family members.

A total of 16 tools provide health and wellness information, which encompasses a variety of subcategories (see Table I). Five tools report on *diet and* *fitness*, returning results such as carbohydrate sensitivity, antioxidant needs, metabolic efficiency, aerobic potential, and injury risk. Eight tools return *pharmacogenetics* information, namely genetic effects on medication response, dosage, and side effects. Two of these tools, Genetic Genie and NutraHacker, provide “detoxification” reports primarily consisting of user genotypes at variants in cytochrome P450 (CYP) genes. These two also provide *methylation* reports: variations in *MTHFR* and other methylation pathway genes. Twelve tools return results on various *traits:* personality and/or physical features. For example, DNA.land’s trait section includes eye color and educational attainment, though developers note additional trait reports are forthcoming (Aufrichtig & Yuan, 2016). DNA Doctor includes a “mental make-up” section with results on novelty seeking and anxiety, while GENETIConcept advertises a “well-being” section on neurotransmitters and brain function. Two tools (DNA Doctor and NutraHacker) also provide *carrier status* results. *Complex disease* information is provided by nine tools, ranging from information on a few conditions to several thousand.

A caveat to the above subcategorizations is that the range of health and wellness results provided by some tools (notably Promethease, openSNP, LiveWello, and Golden Helix Genome Browser) is determined by the databases that populate their reports (see next section), making it challenging to assign available results into predefined subcategories. To illustrate, as entries are added to the crowd-sourced annotation wiki SNPedia, so will the information available from Promethease and openSNP expand. LiveWello’s reports are similarly dynamic due to the “SNP Sandbox” feature, where users can create and share gene report templates with other users, incorporating any genotyped SNP with an assigned reference SNP ID (rsID) (LiveWello, n.d.).

Moreover, there is a broad range of ways in which tools present information across these health and wellness subcategories, largely due to the type of analysis performed. For example, several tools just display user genotypes alongside SNP-level information extracted from publicly-available variant annotation or publication databases, as noted above, with minimal additional contextualization. Other tools go beyond these database linking activities, further processing user data to generate an individualized risk estimate, trait score, and/or recommended course of action, discussed further in the following section.

*Analysis of user data: types and sources*

Next, we discuss the types of analyses tools employ to generate users’ results, as well as the information sources used. In Table I, we annotate whether analysis types are peer-reviewed, proprietary, or “homegrown” (i.e., developer’s methodology not clearly marked as proprietary). For tools that return health and wellness results, we further indicate whether the analysis involves aggregating across SNPs, contextualizing against other users’ data, and/or making recommendations. In the table, “database linking” is used to describe analyses that simply link user genotypes to external information sources without further processing or curation.

Tools that analyze genetic ancestry typically adapt existing, *peer-reviewed* methodology (e.g., principal component or admixture analysis) to generate reports, though some companies are founded on novel and *proprietary* algorithms — for example, GPSOrigins (Elhaik et al., 2014). One analytical feature that distinguishes ancestry tools, and by which they often wish to be distinguished, is the *proprietary reference panel* used. For example, WeGene has attracted many international customers with their collection of Asian ancestry reference samples. Tools’ proprietary reference panels often include data from some *public sources*, such as the 1000 Genomes Project and the Human Genome Diversity Project. DNA Tribes and Family Tree DNA also incorporate customer data into their reference panels.

Genealogy-focused tools rely on standard, *peer-reviewed* methods for estimating relatedness (e.g., identity by descent analyses) or on *homegrown* tools written by the developer(s) (i.e., David Pike’s utilities and James Lick mtDNA haplogroup analysis). GEDMatch appears to use a proprietary relative finder algorithm (“GEDMatch Autosomal comparison software”) and an assortment of ancestry calculators created by different “citizen scientists.” A key difference between genealogy analyses across tools is the search space for finding relatives: tools that analyze data locally (e.g., David Pike utilities) can only evaluate data files the user has collected (e.g., obtained from other individuals). DNA.land and GEDMatch, on the other hand, centrally store data and therefore can search for relatives across the full user database.

All tools returning health and wellness information rely to some extent on scientific publications and/or publicly-available variant annotation databases, such as ClinVar (Landrum et al., 2016), dbSNP (Sherry et al., 2001), and the NHGRI GWAS catalog (Welter et al., 2014). Where the tools differ analytically is the extent to which they further curate or process these reference sources, such as through in-house *literature review*. At one extreme, tools simply link users to external information sources (described as *database linking* in Table I). For example, Promethease displays user genotypes alongside entries from the crowd-sourced variant annotation wiki SNPedia — the developers’ companion project, which is populated from a combination of manual and automated annotation. The results of these database linking activities are typically presented SNP by SNP, showing user genotypes at the selected variant against information extracted from the external database, for potentially thousands of diseases or traits.

Tools go beyond database linking by doing one or more of the following activities: aggregation, contextualization, and/or recommendation. *Aggregation* across variants involves selecting variants to aggregate and selecting an algorithm or formula for performing the aggregation (e.g., weighting SNP-level effect sizes to generate a genetic risk score, or GRS). For example, some DNA Doctor reports generate an overall risk by summing SNP-level risks extracted from SNPedia (low, medium, high), where the set of SNPs has been selected by the tool developer. Interpretome uses a slightly more sophisticated aggregation approach: calculating a GRS using a formula and SNP list from selected scientific publications. *Contextualization* means presenting the user’s results in the context of a given population (e.g., other users or a public reference such as HapMap). For instance, DNA.land and GeneKnot both present the user’s trait score or risk estimate in a histogram populated with other user data. A third way tools go beyond database linking is by making *recommendations*: offering advice or suggested actions based on interpretation of the user’s genetic profile. The tools making the most recommendations are those returning diet and fitness reports: AnabolicGenes, Athletigen, DNAFit, GENETIConcept, and NutraHacker. The additional curation effort required to aggregate, contextualize, and/or make recommendations leads to these tools typically returning results on fewer traits or conditions as compared to database linking tools.

***Interviews with developers***

Tools are created by a diverse set of individuals and entities: non-specialists/citizen scientists (3 tools), commercial companies (15 tools), and academics from genetics or related fields (e.g., bioinformatics — 5 tools). The category “Company” includes some small Limited Liability Companies (LLCs; e.g. Genetic Genie and DNA Doctor); not all companies charge a fee for reports. Interview participants represented all three categories: one non-specialist/citizen science developer, three developers from commercial companies, and six developers representing four academic tools. Here, we briefly describe several high-level themes from the overall interview data, then focus on a subset of results related to developers’ views on the purview of their tools.

Interviews with tool developers helped contextualize the content analysis results by illuminating developers’ rationales and motivations for how tools generate and present information to users. Several developers, for example, described building the tool they wanted for themselves. Specifically, they were personally unsatisfied with existing platforms for interpretation and/or sharing their own genetic data, including those provided by the DTC-GT companies, and therefore designed a platform to fill the perceived gap. Developers recognized that users vary in their level of expertise or familiarity with genomics, and many assumed the tool interface would accommodate this range by allowing a self-selecting group of users to delve deeper into linked resources. Developers also held complex views of the medical community. Some were frustrated with a perceived underutilization of genomics in health care and saw their tool as a way to allow users to circumvent this. Developers of one tool, however, have instead cultivated a relationship with genetic counselors, with whom they described having a shared mission of increasing genetic literacy in the general public. Example quotes from these themes not discussed in more detail below are presented in Online Resource 2.

Notably, several tool developers (5, or 50% of those interviewed) appeared to challenge or reject the idea that their tools perform “interpretation.” Instead they characterized tools as a connection to existing scientific literature repositories or annotation databases (e.g., PubMed, ClinVar, SNPedia), acting as what one developer called a “bridge to the literature.” By characterizing their activities this way, developers seem to distance themselves from entities — including the major DTC-GT companies — that provide more personalized, “packaged” information. Specifically, respondents articulated a difference between simply linking genotypes to literature or annotation databases at the level of individual variants (bridging) versus aggregating information across multiple variants to create some type of personalized report or risk assessment (interpreting). Note this distinction was made primarily in the context of providing information related to health and wellness, rather than ancestry or genealogy. The reasons developers gave for bridging versus interpreting fell into three domains: scientific, ethical, and regulatory. Table II presents a summary of explanations for bridging and the supporting quotes presented below.

*Explanations for bridging: scientific*

One major justification for bridging was based in an understanding of genetics as a developing field, where the current state of knowledge about genotype-phenotype associations makes it premature to go beyond bridging. By linking users directly to resources such as SNPedia, for example, rather than attempting to further combine or package the information, one developer viewed his tool as more faithfully representing the complexity of the state of the science. “We sort of avoid a high gloss, glitzy, eye candy simplistic view of things, we're really trying to stick very tightly to: this is a bridge to the scientific literature, and if the literature is complex — well, then, your report's going to be complex” (ID 1336).

Some developers commented specifically on the limitations of combining information across SNPs to produce a single score or estimate of disease risk. One developer felt it was premature to try and provide users with a single risk estimate given the small proportion of heritability explained for most complex diseases: “I don't think the time is yet ripe for that because I don't think the field has matured yet to the point…[where you can] give people a [single] number” (ID 1077). Another observed that while polygenic risk scores are mathematically sound, “ [there are] no prospective or even retrospective studies on SNP profiles to outcomes…the problem is…we have not validated that” (ID 1005).

*Explanations for bridging: ethical*

Another class of justifications for bridging was centered in ethical considerations, including enhancing transparency, educating, and avoiding coercion to contribute data. These ethical commitments often stemmed from recognition of the complexities and uncertainties of genetics, similar to the justifications discussed in the previous section. However, the ethical concerns depart from the scientific ones above in that developers reference downstream harms that might arise from providing users with information unsupported by the science. “Ethically, we are certainly…aligned with the groups who feel it’s improper to do things like summarize SNPs in ways the scientific literature does not offer any evidence to support. And certain companies have done that in the past, or even continue to do it today” (ID 1336). Therefore, rather than oversimplify or obscure complexity, this developer seemed to value transparently presenting the complexity in full to users via linking to primary literature. Another developer’s comments expand on this idea: that linking to literature is not only desirable in and of itself, as a faithful representation of the science, but furthermore empowers users with an “option to go deeper.” He elaborated that by linking users’ reports to SNPedia, “those who want to read the study that was associated with that SNP can go and do so and decide whether it fits — if it holds up to their standard of what a good study looks like” (ID 1077).

Bridging to the literature is also valued as a means of educating users, with the goal of increasing overall genetic literacy. One tool, for example, Interpretome, began as a series of classroom exercises to teach medical students about genetics (Karczewski et al., 2012) and is described by its developers as a “teaching tool” (ID 1782). Specifically, it is meant to show users how DTC-GT companies generate their reports: “The point of the website is really not to do any interpretation, it's to show you how you would do interpretation. It is really, you know, it is an educational tool” (ID 1005).

A slightly different ethical motivation for bridging exists for the tool openSNP, which makes user genotype data freely and publicly available and therefore takes steps to ensure that users weigh risks and benefits carefully prior to use. “I always tend to say that we try and scare away people actively. By putting out like, all the negative things that like might happen even if there is no evidence at all that this has really happened…it's for this reason that we also don't try to do too much analysis to keep people from not thinking there's like a big benefit to openly sharing it, but rather you probably will not get any benefit out of it” (ID 1671).

*Explanations for bridging: regulatory*

Finally, a third set of reasons to bridge was articulated relative to regulatory concerns. Some developers were uncertain about whether they could be subject to existing regulatory frameworks, and to mitigate such uncertainty reported erring on the side of caution in at least one of two respects: (1) the categories of information returned (i.e., limiting to “wellness” rather than health-related traits) and/or (2) in the way information is presented. As one developer noted, “they [FDA] don't have a policy and they're not in a hurry to develop a policy. And this is like the worst type of regulation, because if there is a policy at least I know what I can and cannot do. When there is no policy…you have to guess…So, right now we decide to go the safe setting not to give any health information. Just fun traits…well-being, physical traits” (ID 1702). Another developer similarly described having “good reason to keep it in the wellness arena, just to sort of be clear liability-wise with the FDA” (ID 1077).

Apart from limiting the categories of results returned, some developers also hedged regulatory uncertainty by trying to formulate the tool as, again, a bridge to the literature. Here developers portrayed bridging (versus interpretation) as less likely to receive regulatory scrutiny, namely that linking individuals to publications or annotation database is distinct from providing health-related interpretation or advice. For example, one developer described why presenting user genotypes next

to variant records in SNPedia, PubMed, dbSNP, and the GWAS catalog does not amount to interpretation: “here is your SNPs and click here if you want to read a report about this SNP in general, right?...I don't see how it is an interpretation because we just, you know, we basically just put one thing next to each other, we don't fuse them, it's up to the user to fuse if he or she wants to do that” (ID 1702). Another developer described their tool as “too low level” to fall under FDA jurisdiction; that it was “kind of a stretch to say that [the tool] is somehow giving advice” because while supporting literature is presented, no report is being generated from it (ID 1671).

**Discussion**

We characterized and analyzed 23 third-party interpretation tools for raw/uninterpreted personal genetic data from DTC-GT, via structured content analysis of tool websites and qualitative interviews with a subset of tool developers. Third-party tools vary considerably with respect to the types of information returned to users, the analytic and bioinformatic approaches used to generate the information, and the level of transparency to users about analysis and interpretation. In interviews with tool developers, we gained insight into the motivations and rationales driving tool design and, in particular, choices of information to return to users. Notably, for many of the tools providing health-related and wellness information, developers challenge the claim that their tool interprets genetic data. Instead, tools are viewed as simply “bridging” the user to the scientific literature, via linking to publication and variant annotation databases.

While in interviews, several developers emphasized the bridging aspects of their tools, our content analysis results and further examination of example tool reports suggest that instead there is a continuum between bridging and interpretation, and several tools in fact are located along this continuum. This makes sense given that bridging is a stepping stone on the way to interpretation. For example, Promethease reports are generated by linking user genotypes to SNPedia entries and thus exemplify bridging. However, SNPedia entries can include recommended behavior(s). Because the SNPedia entry content varies across variants, and Promethease returns information about thousands of variants, it is difficult to classify the tool overall as purely bridging versus an intermediate between bridging and interpretation. Developers themselves may not have a clear understanding of whether their tool is primarily bridging or interpreting. For example, one of the developers that described theirs as a “database tool” performs aggregation across variants to produce an overall risk. In summary, we have identified a novel distinction between bridging versus interpretation among third-party tools; however, our understanding of these distinctions, and the continuum between them, is likely to evolve as the number and nature of third-party tools and consumer genetics more broadly continues to expand.

***Practice Implications***

Providers have been grappling with patients’ DTC-GT genetic findings for years (Brett, Metcalfe, Amor, & Halliday, 2012; McGowan, Fishman, Settersten, Lambrix, & Juengst, 2014; Powell, Christianson, et al., 2012; Powell, Cogswell, et al., 2012; van der Wouden et al., 2016); however, the scale and complexity of information that may be conveyed from third-party interpretation of raw genetic data is unprecedented and likely to pose significant new challenges for health care providers (Allen, Gabriel, Flynn, Cunningham, & Wang, 2017; Borry et al., 2017). Specialists such as genetic counselors and medical geneticists are particularly likely to encounter patients with third-party reports. As Kirkpatrick and Rashkin (2017) observe, in the face of growing access to raw genetic data and third-party interpretation services, “the role of the genetic counselor is likely to evolve dramatically.” DTC-GT companies’ predicted shift from SNP array genotyping to whole exome or genome sequencing will likely contribute to the genetic counselor’s expanding role. Specifically, while uninterpreted data files from most DTC-GT companies currently include less than 1 million SNPs assayed by array genotyping, a progression to sequencing technology would enable customers to download their genotypes at potentially millions of variants. The scale of “raw” data available will coincide with an increased scale of information individuals may seek from third-party tools and other online information sources. Providers, both inside and outside genetic specialties, may be expected to act as a “buffer” between patients and the glut of genetics and other health-related information available online (Murphy, 2009). Of course, it is not clear that providers can or should perform this buffering role, in the face of competing demands on time and resources.

While tool developers may value faithfully representing the complexity of genetics, it is unclear how these “bridging” activities are experienced by users and the health professionals they are likely to consult. One possibility is that the lack of contextualization or explanation of results will confuse and frustrate users and ultimately not deter them from perceiving the reports as medical advice (Badalato et al., 2017). Alternatively, bridging users to scientific literature may empower them to “dig deeper” at their discretion, promoting autonomy and transparency as intended by several developers we interviewed. Furthermore, the lack of additional processing may make it easier for providers to highlight the limitations of the source information, as compared to third-party tools that aggregate or otherwise potentially obscure research findings. Indeed, a recent ethical analysis of third-party interpretation recognized that, while difficult to parse, non-aggregated SNP-level information may ultimately give users a “more realistic perspective of the uncertain nature of multifactorial disease prediction” (Badalato et al., 2017). Our results extend that analysis by revealing developers’ rationales for bridging, which though partly driven by ethical considerations such as transparency and promoting autonomy, also stem from regulatory concerns.

 The complications of when and how to adjudicate third-party interpretations are only likely to intensify for genetics professionals as access to raw data, and available third-party tools, expands. While historically DTC-GT has been the most common route of access, there are at least two major additional ways in which members of the public will be able to access their raw personal genetic data moving forward: (1) through clinical sequencing tests and the HIPAA direct access right (45 C.F.R. § 164.524) and (2) via participation in research studies that offer raw genetic data to participants. As the supply of raw personal genetic data expands, a growing number of third-party tools are likely to crop up to meet demand. For example, since freezing the dataset for our current study, we have become aware of additional tools, such as Self Decode ([www.selfdecode.com](http://www.selfdecode.com)) and CodeGenEU (<https://codegen.eu>).

***Study Limitations***

 Because the range of third-party tools can be daunting, we intend our investigations to facilitate providers’ decisions about how and when to respond to patients’ third-party reports. However, the information about third-party tools available on their websites is limited and indeed in some cases biased, particularly for proprietary systems. We sought interviews with tool developers to extend and augment information gleaned from third-party tool websites, although not all developers agreed to be interviewed. While we were able to interview three developers of commercial tools, our interviewees were biased towards developers from academia, likely reflecting overall familiarity with and willingness to participate in academic research, in addition to having fewer proprietorial concerns. In addition, while we have described the types of reports available across tools in this study, it is difficult to give an exhaustive and detailed list of all available reports in an easily digestible manner (i.e., in Table I). Available reports may have also changed, either expanded or reduced, since the time of study.

***Research Recommendations***

 We have reported on a novel area of user-driven interpretation of personal genetic data via third-party tools. This work is an important albeit preliminary step to further understanding of personal data access across commercial, research, and clinical realms. Currently very little is known about the *users* of third-party tools, including how they perceive and digest the information, and when and how they might try to integrate the information with their health care. Uncertainty also exists about the potential for regulation of third-party, interpretation-only tools (Evans, 2014; Lucivero & Prainsack, 2015; Spector-Bagdady & Pike, 2013). While a full discussion of these regulatory complexities is outside the scope of the present study, we note that in our interviews, several tool developers expressed concern and uncertainty regarding whether and how they might be regulated (e.g., see example quotes in Online Resource 2). Regulatory concerns are further complicated by jurisdictional issues that likely arise for web-based services operating across state and national boundaries. Future work should interrogate both the perspectives and experiences of the DTC-GT consumers who are using these third-party tools and the regulatory environments in which third-party tools are operating.

**Acknowledgements**

We thank the interview participants as well as the organizers and attendees of the 2017 “Genomics and Society: Expanding the ELSI Universe” Congress, where an earlier version of this work was presented. We are also grateful to Deborah J. Bowen and Barbara J. Evans for early input on study design and scope. This work was supported by the National Human Genome Research Institute (NHGRI) and the National Cancer Institute (NCI) Clinical Sequencing Exploratory Research (CSER) Consortium, U01HG006507 and U01HG007307 (Jarvik, PI). This work was conducted as part of Sarah C. Nelson’s doctoral dissertation in Public Health Genetics at the University of Washington.

**Conflict of Interest**

Sarah C. Nelson and Stephanie M. Fullerton declare that they have no conflict of interest.

**Human Studies and Informed Consent**

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. This study was approved by the University of Washington Institutional Review Board as minimal risk human subjects research (approval #50238). Informed consent was obtained from all individual participants included in the study.

**References**

Allen, C., Gabriel, J., Flynn, M., Cunningham, T. N., & Wang, C. (2017). The impact of raw DNA availability and corresponding online interpretation services: a mixed methods study. *Translational Behavioral Medicine*.

Aufrichtig, R., & Yuan, J. (2016, October 14). DNA.Land’s Trait Prediction Report. *Medium (Weblog)*. Retrieved from https://medium.com/@dl1dl1/dna-lands-trait-prediction-report-8efbd3a74f98

Autosomal DNA tools. (n.d.). Retrieved August 12, 2017, from https://isogg.org/wiki/Autosomal\_DNA\_tools

Badalato, L., Kalokairinou, L., & Borry, P. (2017). Third party interpretation of raw genetic data: an ethical exploration. *European Journal of Human Genetics*.

Bettinger, B. (2013). What Else Can I Do With My DNA Test Results? Retrieved November 10, 2014, from http://www.thegeneticgenealogist.com/2013/09/22/what-else-can-i-do-with-my-dna-test-results/

Bloss, C. S., Wineinger, N. E., Darst, B. F., Schork, N. J., & Topol, E. J. (2013). Impact of direct-to-consumer genomic testing at long term follow-up. *Journal of Medical Genetics*, *50*(6), 393–400.

Bobe, J. (n.d.). Sharing genome studies. Retrieved May 14, 2017, from http://blog.jasonbobe.net/sharing-genome-studies/

Borry, P., Bentzen, H. B., Budin-Ljøsne, I., Cornel, M. C., Howard, H. C., Feeney, O., … Felzmann, H. (2017). The challenges of the expanded availability of genomic information: an agenda-setting paper. *Journal of Community Genetics*, 1–14.

Brett, G. R., Metcalfe, S. A., Amor, D. J., & Halliday, J. L. (2012). An exploration of genetic health professionals’ experience with direct-to-consumer genetic testing in their clinical practice. *European Journal of Human Genetics*, *20*(8), 825–830.

DNA Testing Reviews. (n.d.). Retrieved August 12, 2017, from https://dnatestingchoice.com/

Elhaik, E., Tatarinova, T., Chebotarev, D., Piras, I. S., Maria Calò, C., De Montis, A., … Ziegle, J. S. (2014). Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nature Communications*, *5*.

Erlich, Y., Gordon, A., Pearson, N., Shee, K., & Pickrell, J. (2015). DNA.Land: A community-wide platform to collect millions of genomes-phenomes. Presented at the 65th Annual Meeting of The American Society of Human Genetics. Baltimore, MD.

Evans, B. J. (2014). The First Amendment Right to Speak about the Human Genome. *University of Pennsylvania Journal of Constitutional Law*, *16*(3), 549–636.

Evans, B. J., Dorschner, M. O., Burke, W., & Jarvik, G. P. (2014). Regulatory changes raise troubling questions for genomic testing. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *16*, 799–803.

Global Alliance for Genomics and Health. (n.d.). Participant Values Task Team. Retrieved August 27, 2017, from https://genomicsandhealth.org/working-groups/our-work/participant-values

Greshake, B., Bayer, P. E., Rausch, H., & Reda, J. (2014). openSNP-A Crowdsourced Web Resource for Personal Genomics. *PLoS One*, *9*(3), e89204.

Karczewski, K. J., Tirrell, R. P., Cordero, P., Tatonetti, N. P., Dudley, J. T., Salari, K., … Kim, S. K. (2012). Interpretome: a freely available, modular, and secure personal genome interpretation engine. *Biocomputing*, 339–350.

Kaufman, D. J., Bollinger, J. M., Dvoskin, R. L., & Scott, J. A. (2012). Risky business: risk perception and the use of medical services among customers of DTC personal genetic testing. *Journal of Genetic Counseling*, *21*(3), 413–22.

Kirkpatrick, B. E., & Rashkin, M. D. (2017). Ancestry Testing and the Practice of Genetic Counseling. *Journal of Genetic Counseling*, *26*(1), 6–20.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., … Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, *44*(D1), D862–D868.

LiveWello. (n.d.). Genetics tools. Retrieved January 6, 2018, from https://livewello.com/genetics

Lucivero, F., & Prainsack, B. (2015). The lifestylisation of healthcare? “Consumer genomics” and mobile health as technologies for healthy lifestyle. *Applied and Translational Genomics*, *4*.

Lunshof, J., Church, G., & Prainsack, B. (2014). Raw Personal Data: Providing Access. *Science*, *343*(6169), 373–374.

McGowan, M. L., Fishman, J. R., Settersten, R. A., Lambrix, M. A., & Juengst, E. T. (2014). Gatekeepers or intermediaries? The role of clinicians in commercial genomic testing. *PloS One*, *9*(9), e108484.

Murphy, S. (2009). In need of a reality check. *Nature Biotechnology*, *27*(5), 422–422.

Nelson, S. (2016). Geneticists should offer data to participants. *Nature*, *539*(7627), 7–7.

Open Humans. (n.d.). Retrieved August 12, 2017, from www.openhumans.org

Powell, K. P., Christianson, C. A., Cogswell, W. A., Dave, G., Verma, A., Eubanks, S., & Henrich, V. C. (2012). Educational Needs of Primary Care Physicians Regarding Direct-to-Consumer Genetic Testing. *Journal of Genetic Counseling*, *21*(3), 469–478.

Powell, K. P., Cogswell, W. A., Christianson, C. A., Dave, G., Verma, A., Eubanks, S., & Henrich, V. C. (2012). Primary care physicians’ awareness, experience and opinions of direct-to-consumer genetic testing. *Journal of Genetic Counseling*, *21*(1), 113–126.

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311.

Spector-Bagdady, K., & Pike, E. (2013). Consuming Genomics: Regulating Direct-to-Consumer Genetic and Genomic Information. *Nebraska Law Review*, *92*(4), 677–745.

Sterling, R. (2008). The on-line promotion and sale of nutrigenomic services. *Genetics in Medicine*, *10*(11), 784–796.

The Precision Medicine Initiative NIH. (2017). Return of Genetic Results in the All of Us Research Program. Retrieved August 27, 2017, from https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/return-of-results-agenda-0306-0717.pdf

U.S. DHHS. (2016). Individuals’ right under HIPAA to access their health information 45 CFR § 164.524. Retrieved September 21, 2017, from http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html

van der Wouden, C. H., Carere, D. A., Maitland-van der Zee, A. H., Ruffin, M. T., Roberts, J. S., & Green, R. C. (2016). Consumer Perceptions of Interactions With Primary Care Providers After Direct-to-Consumer Personal Genomic Testing. *Annals of Internal Medicine*, *164*(8), 513–22.

Wagner, J. K., Cooper, J. D., Sterling, R., & Royal, C. D. (2012). Tilting at windmills no longer: a data-driven discussion of DTC DNA ancestry tests. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *14*(6), 586–93.

Wang, C., Cahill, T. J., Parlato, A., Wertz, B., Zhong, Q., Cunningham, T. N., & Cummings, J. J. (2017). Consumer use and response to online third-party raw DNA interpretation services. *Molecular Genetics & Genomic Medicine*, (July), 1–9.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., … Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, *42*(Database issue), D1001-6.

Table I. Content summary of third-party interpretation tools.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **User Data** | **Type of Results** | **Analysis** |  |
| **Name** | **Input formatsa** | **Retention and sharing** | **Genetic ancestry** | **Genealogy and relatives** | **Health and wellness** | **Types** | **Sources** | **Developer type** |
| AnabolicGenes | GDF | Retained; may be shared |  - |  - | Diet and fitness | Proprietary; aggregates and makes recommendations | Lit review | Company |
| Athletigen | GDF; 23andMe API | Retained; may be shared |  - |  - | Diet and fitness; traits | Proprietary; aggregates and makes recommendations | Lit review | Company |
| David Pike's utilities | GDF | None (analyzed locally) |  - | Homozygosity; haplotype phasing; shared segments | - | Homegrown | None (self-contained) | Non-specialist/citizen scientist |
| DNA Doctor | GDF; 23andMe API | Not retained (data from API analyzed locally) |  - |  - | Diet and fitness; pharmacogenetics; traits; carrier status | Database linking; aggregates | SNPedia | Company |
| DNAFit | 23andMe API | Retained; may be shared |  - |  - | Diet and fitness | Proprietary; aggregates and makes recommendations | Lit review | Company |
| DNA.landb | GDF | Retained; may be shared | Global  | Relative finder; shared segments | Traits | Peer-reviewed; aggregates and contextualizes | Lit review; both public and private sources for ancestry and imputation  | Academic |
| DNA Tribes | GDF | Retained; sharing unclear | Global  |  - | - | Proprietary | Proprietary reference panel | Company |
| Enlis Genome Personal | GDF; VCF | Not retained (software product analyzes locally) |  - | Track variants across family members | Traits; complex diseases | Proprietary | PubMed; privately curated databases | Company |
| Family Tree DNA (Autosomal Transfer) | GDF | Retained; may be shared | Global  | Relative finder; shared segments | - | Proprietary (relative finder); peer-reviewed (ancestry) | Proprietary reference panel  | Company |
| GEDMatch | GDF; 23andMe API | Retained; user controls sharing  | Global; local  | Relative finder; shared segments | - | Proprietary | Proprietary reference panel | Company |
| GeneKnot | GDF | Not retained |  - |  - | Traits; complex disease | Database linking; homegrown; aggregates and contextualizes | Lit review; GWAS Catalog | Non-specialist/citizen scientist |
| Genetic Genie | GDF; 23andMe API | Not retained |  - |  - | Pharmacogenetics; methylation | Database linking | Unclear | Company |
| GENETIConcept | GDF | Retained; may be shared  |  - |  - | Diet and fitness; pharmacogenetics; traits; complex disease | Unclear; aggregates and makes recommendations | Unclear | Company |
| Golden Helix Genome Browser | GDF; VCF; BAM; FASTA | None (analyzed locally) |  - |  - | Traits; complex disease | Database linking | Public sources: e.g., dbSNP, ClinVar, GWAS Catalog, dbNSFP, and SIFT | Company |
| GPS Origins | GDF | Retained; may be shared  | Global; migration maps |  - | - | Proprietary | Proprietary reference panel, includes public sources | Company |
| Infinome | 23andMe API | Retained; only used internally |  - |  - | Traits; complex disease | Database linking | GWAS Catalog | Academic |
| Interpretome | GDF | None (analyzed locally) | PCA; local  |  - | Pharmacogenetics; traits; complex disease | Peer-reviewed (PCA); homegrown (local ancestry); database linking; aggregates and contextualizes | GWAS Catalog; PharmKGB; PolyPhen; and Lit review. Reference panel includes public and private sources | Academic |
| James Lick Haplogroup Analysis | GDF; FASTA, GenBank, ASN1 | Not retained | mtDNA haplogroup |  - | - | Homegrown | PhyloTree | Non-specialist/citizen scientist |
| Livewello | GDF; VCF | Retained; user controls sharing |  - |  - | Pharmacogenetics; traits; complex disease | Proprietary; appears primarily database linking | Lit review; GWAS Catalog; dbSNP | Company |
| NutraHacker | GDF; 23andMe API | Retained; sharing unclear |  - |  - | Pharmacogenetics; methylation; carrier status | Unclear; makes recommendations | Lit review | Company |
| openSNP | GDF; VCF | Retained; publicly available |  - |  - | Pharmacogenetics; traits; complex disease | Database linking | GWAS Catalog; SNPedia; Mendeley; GET Evidence System; PLoS | Academic |
| Promethease | GDF; VCF; 23andMe API | Not retained |  - |  - | Pharmacogenetics; traits; complex disease | Database linking | SNPedia  | Academic |
| WeGene (English version) | GDF; 23andMe API | Retained; may be shared | Global; mtDNA and chrY haplogroup | - | - | Peer-reviewed | Proprietary reference panel, includes public sources | Company |

a – GDF= genotype data file from one or more DTC-GT company; API=application program interface; VCF=variant call format; ASN1=Abstract Syntax Notation One.

b – Primary DNA.land tool only. A companion tool, DNA Compass (<http://compass.dna.land/>), locally analyzes a VCF of imputed data provided by DNA.land. DNA Compass links to variant-level health and wellness-related databases.

Table II. Quotes from third-party tool developers on the theme of “bridge to the literature.”

|  |  |
| --- | --- |
| **Explanations for “bridging”** | **Example quote(s)** |
| Scientific | “We sort of avoid a high gloss, glitzy, eye candy simplistic view of things, we're really trying to stick very tightly to: this is a bridge to the scientific literature, and if the literature is complex — well, then, your report's going to be complex.” (1336)“I have a certain roadmap for the application that includes giving more complexity in terms of maybe a number. Like, ‘here's your risk.’ But I don't think the time is yet ripe for that because I don't think the field has matured yet to the point, you know when we can say we've accounted for 95% of the heritable difference of a trait, then I think the time is ripe to give people a number.” (1077)“[There are] certainly no prospective or even retrospective studies on SNP profiles to outcomes or anything like that. There's more and more work these days on polygenic risk scores…the problem is…we have not validated that.” (1005) |
| Ethical | “Ethically, we are certainly…aligned with the groups who feel it's improper to do things like summarize SNPs in ways the scientific literature does not offer any evidence to support. And certain companies have done that in the past, or even continue to do it today.” (1336)“Basically this is just a database tool is what I'm providing…I want to leave people with an option to go deeper, which I think that's what I've done with the linking each SNP to SNPedia. So that those who want to sort of read the study that was associated with that SNP can go and do so and decide whether it fits — if it holds up to their standard of what a good study looks like.” (1077)[Describing tool report] “I cut and pasted the histogram [from the paper] and I decided I want to look up, for the students, I want to look up their score so they can see where they are on the histogram…And there is a teaching tool. So, in this case it's also effective because if you're pretty far off from the middle it means something for you.” (1782)“The point of the website is really not to do any interpretation, it's to show you how you would do interpretation. It is really, you know, it is an educational tool.” (1005)“I always tend to say that we try and scare away people actively. By putting out like, all the negative things that like might happen even if there is no evidence at all that this has really happened…it's for this reason that we also don't try to do too much analysis to keep people from not thinking there's like a big benefit to openly sharing it, but rather you probably will not get any benefit out of it.” (1671) |
| Regulatory | “They [FDA] don't have a policy and they're not in a hurry to develop a policy. And this is like the worst type of regulation, because if there is a policy at least I know what I can and cannot do. When there is no policy…you have to guess…And you don't get clear answers. We asked them, why Promethease is ok and 23andMe is not ok? They just give you this answer…‘Yeah, it's on our radar, but we really don't have a policy right now.’ So, how can we decide what to do next? So, right now we decide to go the safe setting not to give any health information. Just fun traits that, you know, well-being, physical traits.” (1702)“I definitely didn't want to involve…FDA, so that was another good reason to keep it in the wellness arena, just to sort of be clear liability-wise with the FDA.” (1077)[describing tool report] “Here is your SNPs and click here if you want to read a report about this SNP in general, right? And read about what each genotype means…I don't see how it is an interpretation because we basically just put two things, one thing next to each other, we don't fuse them, it's up to the user to fuse if he or she wants to do that.” (1702)“I personally think it's kind of a stretch to say that this is somehow giving advice…While we do have the association between the different genotypes and the associations found so far, we are not calculating a report out of it…You can go to each individual variation and look up what's your personal variation and what literature says about it…in that case, even each individual primary publication is somehow then informing people…I think we are a bit too low level, let's say, to actually somehow fall under it [regulations for returning health information].” (1671) |

Online Resource 1. Supplementary table of third-party interpretation tool information.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | URLa | Start Year | Offers genotyping service |
| AnabolicGenes | [anabolicgenes.com](https://anabolicgenes.com/) | 2015 | Yes |
| Athletigen | [athletigen.com](https://www.athletigen.com/) | 2014 | Yes |
| David Pike's utilities | [math.mun.ca/~dapike/FF23utils](http://www.math.mun.ca/~dapike/FF23utils) | 2010 |  |
| DNA Doctor | [biostatushealth.com/dnadoctor](http://biostatushealth.com/dnadoctor/) | 2015 |  |
| DNAFit | [dnafit.com](https://www.dnafit.com) | 2013 | Yes |
| DNA.land | [dna.land](https://dna.land) | 2015 |  |
| DNA Tribes | [dnatribes-snp.com](https://dnatribes-snp.com) | 2006 | Yes |
| Enlis Genome Personal | [enlis.com/personal\_edition.html](https://www.enlis.com/personal_edition.html)  | 2015 |  |
| Family Tree DNA (Autosomal Transfer) | [familytreedna.com/AutosomalTransfer](https://www.familytreedna.com/AutosomalTransfer) | 2014 | Yes |
| GEDMatch | [gedmatch.com](https://www.gedmatch.com) | 2011 |  |
| GeneKnot | [geneknot.com](https://geneknot.com) | 2013 |  |
| Genetic Genie | [geneticgenie.org](http://geneticgenie.org) | 2013 |  |
| GENETIConcept | [geneticoncept.com/23andme.html](https://geneticoncept.com/23andme.html) | 2016 | Yes |
| Golden Helix Genome Browser | [goldenhelix.com/products/GenomeBrowse](http://goldenhelix.com/products/GenomeBrowse) | 2012 |  |
| GPS Origins | [homedna.com/gpsorigins](https://homedna.com/gpsorigins)  | 2016 | Yes |
| Infinome | [infino.me](https://www.infino.me) | 2013 |  |
| Interpretomeb | [genotation.stanford.edu](http://genotation.stanford.edu/) | 2011 |  |
| James Lick Haplogroup Analysis | [dna.jameslick.com/mthap](https://dna.jameslick.com/mthap) | 2010 |  |
| Livewello | [livewello.com](https://livewello.com) | 2013 |  |
| NutraHacker | [nutrahacker.com](http://www.nutrahacker.com/) | 2013 |  |
| openSNP | [opensnp.org](https://opensnp.org) | 2011 |  |
| Promethease | [promethease.com](https://promethease.com/) | 2008 |  |
| WeGene (English version) | [wegene.com/en](https://www.wegene.com/en) | 2014 | Yes |

a – Tool website URL. For brevity, “http(s)” and “www” prefixes are not shown.

b – At the time of study, Interpretome was located at [interpretome.com](http://www.interpretome.com); the website was later renamed “Genotation” and subsequently relocated to the website noted here.

Online Resource 2. Supplementary table of additional themes and example quotes from interviews with third-party tool developers.

|  |  |
| --- | --- |
| **Theme** | **Example quote** |
| Build the tool they want for themselves | "When I saw [DTC-GT raw] data files coming out — well, that's just a data file waiting to be input into a program. Just there wasn't a program yet. But me with my programming ability, my motivation for the genealogy, and these data files becoming available — well, that was just like the perfect storm if you want to put it that way. And I was in the right place at the right time to use the skills that I had to answer questions that I had." (1135) |
| Perceptions of tool users | [on fielding emails from users] "It's obvious some of these people have read up on it and have tried to understand, but obviously this is for the public, not all of them have degrees in biology or statistics or anything like that. So they obviously don't — they may or may not have a...background in this. But they're interested and I like seeing those emails" (1005) |
| Relationship with medical community | "We've had private contacts [with genetic counselors] that then evolved into shared screen sessions where we looked together at data and say, ‘Ok, here's what it means to us, what's it mean to you?’ And you know, we learn from each other. So we've had private one-on-one sessions with genetic counselors so that we understand better where each is coming from...And ultimately we're both trying to make sense of what's the right thing to tell somebody." (1336) |
| Views on regulation of third-party tools | "I think when I got into it, there was only very few players in the field, so there was a little bit of concern that maybe I would get a threatening letter from the FDA but...I think they're not as concerned with the third-party interpretation tools as they are with the one who's actually doing the testing." (1077) |
| Views on regulation of DTC-GT companies | "So, after the FDA letter to 23andMe, it became clear to me that FDA was ignorant. I mean they said, what if a woman chooses to have her breasts removed based on the 23andMe report...the whole idea that anybody would do anything, any kind of surgery based on 23andMe is ludicrous. And the whole idea that they came up with BRCA2 illustrates that they must be ignorant about fundamental aspects of human genetics." ( 1782) |