Published in final edited form as:

Cancer Res. 2019 October 15; 79(20): 5140–5145. doi:10.1158/0008-5472.CAN-19-0769.

Value of collaboration among multidomain experts in analysis of high-throughput genomics data

Daoud Meerzaman^{1,2}, Barbara K. Dunn³

¹National Cancer Institute, Center for Biomedical Informatics and Information Technology

³National Cancer Institute, Division of Cancer Prevention

Abstract

The recent explosion and ease of access to large-scale genomics data is intriguing. However, serious obstacles exist to the optimal management of the entire spectrum from data production in the laboratory through bioinformatic analysis to statistical evaluation and ultimately clinical interpretation. Beyond the multitude of technical issues, what stands out the most is the absence of adequate communication among the specialists in these domains. Successful interdisciplinary collaborations along the genomics pipeline extending from laboratory experiments to bioinformatic analyses to clinical application are notable in large scale, well managed projects such as TCGA. However, in certain settings in which the various experts perform their specialized research activities in isolation, the siloed approach to their research contributes to the generation of questionable genomic interpretations. Such situations are particularly concerning when the ultimate endpoint involves genetic/genomic interpretations that are intended for clinical applications. In spite of the fact that clinicians express interest in gaining a better understanding of clinical genomic applications, the lack of communication from upstream experts leaves them with a serious level of discomfort in applying such genomic knowledge to patient care. This discomfort is especially evident among healthcare providers who are not trained as geneticists, in particular primary care physicians. We offer some initiatives that have potential to address this problem, with emphasis on improved and ongoing communication among all the experts in fields constituting a comprehensive genomic “pipeline”, from laboratory to patient.

The explosion and accessibility of technologies to generate massive sets of genomics (and other omics) data carry great appeal. There is clear promise that emerging omics studies will inform basic biological mechanisms. Furthermore, a plethora of clinical genomics and other omics domains is available to enhance patient care, as so elegantly summarized in the Ideas and Opinions piece by Cheng and Solit (1). Genomics research has huge potential benefits for clinical decision-making, enabling refinement of risk assessment as well as diagnostic and therapeutic interventions to promote “precision medicine”. However, judicious interpretation of genomics data is critical to this process (2).

²Corresponding author: meerzamd@mail.nih.gov, 9609 Medical Center Drive, Room 1W 478, Bethesda, Maryland 20892-9787, Phone: 240-276-5205.

“The authors declare no potential conflicts of interest.”

The concerns regarding implementation of genomics research that we present here underlie all phases of research – discovery studies, translational studies, clinical research, clinical application, direct-to-consumer testing – with the clinical application involving “precision medicine”. In this review, we are directing our analysis of these concerns regarding communication toward their impact on clinical applications. However, the underlying issues in communication apply to all other aforementioned areas of genomics research.

The new technology has democratized genomics investigations to the point that virtually any research group with relatively minimal funding and personnel can generate huge quantities of data (3). Yet, not all groups are ideally equipped to select technologies at all stages of the pipeline, from laboratory platforms to algorithms for data analysis to biological interpretation (Figure 1 panels A1-C). Most concerning, the potential for error is compounded when data interpretation is applied in the clinic. In this setting misinterpretation of genomics data may have adverse consequences for the patient or other individuals seeking health-related directives.

Large-scale genomics data must be robust and reproducible if they are to benefit patients. Reproducibility of analyses of massive datasets generated in genomics research can be challenging (4). When it occurs, lack of reproducibility in genomics data is attributable to a number of critical factors, especially technical and bioinformatics variability and statistical issues. The latter include overfitting and systematic biases inherent in the study design (3, 5). Technical and bioinformatics variability and statistical issues leading to irreproducibility are amplified by limitations to the communication among basic scientists, computational biologists and clinicians, among others (6). In other words, experts in these various disciplines speak different “languages”. This weakness is exacerbated by the fact that typical genomics pipelines function like a relay race, in which the baton is passed in a unidirectional manner without sharing information in either a reverse or lateral direction (Figure 1 panels A1-C). The limitations imposed by operations that mimic a relay race also can lead to omission of essential information that enriches the outcomes generated by the basic genomic pipeline. For example, beyond the basic clinical interpretations of genomic findings, a fully annotated description of the clinical attributes of each patient – including age, age at diagnosis, stage at diagnosis, etc. – should be shared in all directions of the pipeline.

The laboratory scientist passes generated data on to the bioinformaticist but they often do not communicate again. Similarly, the clinician receives a genomic interpretation from the bioinformaticist without necessarily understanding the meaning of this interpretation or the limitations to its validity. Meanwhile, upstream the computational expert may have selected algorithms that do not appropriately address the clinical question. Even further upstream, at the initiation of the pipeline, the overall study design is frequently generated by experts in laboratory procedures without critical input from statisticians. A preferred approach would mimic a team effort involving a multi-directional communication among all experts, in the form of a network (Figure 2 panels A1-C).

Examples exist of successful interdisciplinary collaborations along the genomics pipeline aimed at analysis with clinical application in mind. Large genomics databases that employ such interactive approaches exemplify multi-disciplinary input starting from the very

beginning of the pipeline, with articulation of the hypothesis and the study design. The Cancer Genome Atlas (TCGA) (7), for example, a cancer genomics program, led to molecular characterization of more than 20,000 primary cancers and matched normal samples for 33 types of cancer. This massive effort was a collaborative project involving the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) and brought together researchers with diverse expertise from multiple institutions. The overall aim was to discover key cancer-causing genomic alterations. This was accomplished by integrating multi-dimensional analyses which enabled cataloguing of the genomic findings, which were incorporated into a comprehensive atlas comprising cancer genomic profiles. The generated datasets have been made available to researchers who have used these to develop improved diagnostic tools, as well as novel approaches to cancer treatment and prevention. Evidence for collaboration among the various involved disciplines is apparent in the multi-site arrangement of the different steps in the genomics process; each site had an assigned disciplinary “task”. Thus, sites responsible for obtaining tissue connected to biospecimen core resource sites. The latter catalogued and verified the quality and quantity of received samples and passed the clinical data on to the data coordinating center and the molecular analytes for genomic analysis to the genome sequencing centers and so on. The integration of the tasks carried out by the designated sites was optimal, leading in the end to availability of well curated datasets to the wider community of researchers.

Given the success of the TCGA, subsequent genomics projects of comparable size and ambition were initiated. These include the International Cancer Genome Consortium (ICGC) (8) and the The Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) (9). The purpose of the ICGC is to coordinate multiple research projects with a common goal of discovering the genomic changes potentially derived from the >20,000 tumor genomes that are available worldwide. APOLLO, a multi-federal cancer center initiative started in 2016, involves the NCI, the US Department of Veterans Affairs (US VA) and components of the Department of Defense (DOD), including Walter Reed Hospital and the Uniformed Services University of the Health Sciences (USUHS). The goal is to perform such analyses on 8000 annotated human tissue specimens involving a large range of organ sites.

These massive projects exemplify successful implementation of the optimal approach to collaboration among the specialists participating in genomics projects. We should be aware, however, that at the more local level involving smaller research projects, where well designed infrastructures may not be available to oversee such communication, barriers to interaction among disciplines continue to exist. The barrier to communication posed by discordant languages among the specialties ultimately can lead to variability and erroneous conclusions at both the technical as well as the downstream analytic and interpretative levels. The technical concerns include basic study design as well as choice and implementation of laboratory platforms. Basic study design involves selection of the population and assortment of comparator groups, sample type, and other pertinent features. The contribution of these choices to study design are vulnerable to potential biases. For example, bias, particularly when embedded in the study population, can be introduced when comparator groups have inherent differential qualities that are not overtly acknowledged but

may influence the outcome. Technical variability due to laboratory methodologies may involve differences among laboratories in sample preparation and experimental platforms, as well as initial analysis of generated data. A typical example of a laboratory process that may not be understood by the downstream bioinformaticists is the potential of FFPE (formalin fixed paraffin embedded) tissue preparation and DNA fragmentation to distort the mutation call frequency. This occurs, for example, when overexposure to FFPE leads to artifactually introduced mutations or when DNA is “over-fragmented” by excessive sample sonication. The input quantity of DNA can also affect the mutation diversity, as can heterogeneity of the tumor tissue, involving either non-tumor tissue interspersed with tumor or multiple tumor subclones (10). Additional components of the experimental conditions in the wet laboratory can also impact the variants called downstream in the pipeline. Without knowledge of the laboratory methodology, the bioinformaticist has no way of knowing whether detection of a variant has been influenced by technical features such as those just described. In other words, is the variant real, does it represent a false positive, or does its absence reflect a false negative? In addition, technical variability due to tumor tissue heterogeneity may also be manifest in a low mutation frequency (referred to as MAF/minor allele frequency). Is a low MAF real or is it due to low tumor purity which can obscure existing mutations? The impact that laboratory platforms have on the ultimate read are beginning to be addressed as computational biologists are undergoing training in laboratory procedures and reciprocally laboratory scientists are becoming more literate in computational approaches (11).

The data generated by the sequencing platform in the “wet lab” constitute the “raw data”, also known as the fastQ files. FastQ files contain large amounts of sequence reads that are unnecessary for downstream analysis. Thus, data at the stage of fastQ files require quality control (QC), a step that is generally performed by the bioinformaticist. QC involves many steps including removal of adaptors, repetitive sequences (sequences erroneously duplicated during polymerase chain reaction/PCR), and nucleotide and sequence reads deemed to be ambiguous. Failure to clean up the fastQ data in this fashion can lead to multiple erroneous downstream conclusions, including false positives and false negatives. Essentially, lack of attention to or understanding of this critical “clean-up” step can distort the subsequent analysis performed by the bioinformaticists.

Once the data are in the hands of the bioinformaticists, they must select from a variety of algorithmic approaches to perform alignment of the experimental sequence to a reference sequence. In the case of human genomic investigations, the latest version of the human reference genome, also known as the reference assembly, is used; the current version is reference genome build 38. The alignment process involves first conversion of the QC'd raw data into a SAM (sequence alignment map) file, which is complex. This is subsequently converted into a BAM (binary alignment map) file, which is essentially a binary equivalent and compressed version of the SAM file.

The investigator needs to be aware that different algorithms designed for each stage of the bioinformatic pipeline may differ in terms of their impact on the final variant calls. (12, 13). For example, at the stage of alignment the choice of algorithm can make a difference in the quality of alignments produced (14). While differences in aligner performance still exist, ongoing research in this area has resulted in improved consistency. In our hands, for

example, a comparison of three aligner tools (Bowtie2, BWA and NovoAlign) yielded no meaningful differences in alignment. On the other hand, the pairing of each of these three aligner tools with the same downstream mutation caller algorithm has been shown to impact the actual variant calls (manuscript submitted to Nature Biotechnology). Over and above the impact on variant calls attributable to the aligner-caller combination, inconsistencies in variant calling can emerge from the application of different variant callers in and of themselves (10, 12, 13). This inconsistency is due in large part to lack of community consensus regarding variant calling. Newer guidelines have been suggested to ensure high-quality calling of the variants in sequencing data. One approach to quality control of variant calls is to require that the automated variant calling be followed by manual review of aligned read sequences (15). Consistency in approach to analytic pipelines is critical to obtaining reproducible and meaningful genomic interpretations. Only in this way will the inherent limitations of the bioinformatic process be amenable to clinical decision making.

Regarding the subsequent steps, various statistical methods have been applied to analyses of big data outcomes, including some that address the heterogeneity resulting from combining results from different studies (16). Whereas a discussion of each statistical method is out of the scope of this review, the most common statistical concern, that of overfitting, deserves mention. Overfitting due to multiple testing is especially common in studies addressing large amounts of genomics data, where each observation reflects a separate endpoint and erroneous statistical significance can be attributed to the results as a matter of pure chance (3, 5). Several statistical methods can be applied to correct the raw analytic findings and thereby reduce the chance of false positives. Since the non-statistician needs to understand the danger of statistical analysis in the context of multiple testing, a statistician should be on board and consulted on appropriate statistical approaches from initial study design all the way to downstream analyses.

The key elements of a genomic study - technical variability, bioinformatic and statistical analysis - must be carefully implemented. This can only be done if the involved investigators – the laboratory scientist, the computational experts (bioinformaticists), and the study design experts (statisticians, epidemiologists) are engaged in open and understandable communication from the earliest conception of the study throughout the entire analytic pipeline, or better, if scientists are multi-domain experts. It is in this manner that the investigators can ensure that clinically relevant genetic/genomic findings are accurate and reproducible. As illustrated in the multi-dimensional pipeline presented in Figure 2 panels A1-C, communication among these specialists must be dynamic and interactive in all dimensions. Thus, the bioinformaticists and study design experts must have a thorough understanding of the biological questions that are being asked as well as the laboratory platforms that are used. Conversely, the biologists must be familiar with the computational and statistical approaches that will be used to analyze the generated data. Communication among specialties and communities may well be adequate in well-designed and well-implemented projects such as TCGA. Yet, much of the current genomics science is being carried out in the local context involving small, relatively isolated research groups. In this setting, limited communication among the specialties continues to exist, in large part due to the lack of resources and access to expertise.

Inconsistent communication may not only impair the ability to ensure that researchers in all domains understand the technical aspects of each other's work. A less tangible obstacle to communication can evolve out of essential differences in culture among the various scientific specialties. Simply put, their views of what is important in the overall research endeavor may differ, and this can lead to detrimental effects on the quality of the final genomics product. The downstream ramifications of the limited communication are especially concerning when the intended application of the genomics research question is applied in the clinic community.

Many in the medical community are aware of the potential clinical consequences of the problems with communication among disciplines. Many physicians in non-genetic specialties express interest and see value in genetic testing, especially in the current environment with its emphasis on precision medicine. However, the level of genomics literacy among clinicians may not be sufficient to select and interpret appropriate genetic/genomic tests and to understand the potential for error in the steps along the way to the result reported to them. Many healthcare providers, as well as medical students training in genomics/precision medicine courses, have expressed a lack of confidence in using genomic results in clinical decision making (17, 18).

This lack of confidence stems largely from limited exposure to genomics analytic approaches during medical training. One study on genomics training reports that while medical students supported the use of personalized, i.e. precision, medicine, they felt they lacked preparation for applying personal genome test results to clinical care (17). In a survey study, medical students who took a specialized genomics and personalized medicine course almost universally perceived most physicians as not having enough knowledge to interpret results of personal genome tests (19). This dilemma is exacerbated by the current trend of genomic technologies being brought not only to the clinic but also increasingly directly to consumers, in the form of direct-to-consumer (DTC) testing (20). Consumers now rely on primary care health providers to interpret such DTC-obtained genetic results (21, 22). As articulated recently in a *JAMA* editorial (23), the "...substantial knowledge asymmetry between purveyors of precision health technologies and nongeneticist clinicians has created the potential for well-intentioned misapplication of tests to guide prevention, diagnosis, and treatments, with unknown downstream consequences."

To improve genomic literacy at the provider level, input from health professionals trained in genomics (e.g. genomics counselors, nurse genomicists, medical genomicists) will be necessary. The integration of these experts into clinical services will provide both greater depth of clinical expertise and informal teaching of nongenetic clinicians, i.e. learning on the job. In addition, physicians should take advantage of online genomics resources compiled by organizations such as the NHGRI (24) and the World Health Organization (WHO) (25). These websites provide educational materials such as short online training courses delivering an introduction to genomics concepts, various types of genomics abnormalities, and examples of genomics reports from actual patients. Such resources provide non-geneticist healthcare providers with accessible and easily understandable explanations of the various elements involved in interpretation of clinical genetics results.

Additionally, clinical improvements will benefit from expanding the current model for genetic counselors to a genomics level. Experts in this area, called “clinical genomicists” (26), should be embedded in the clinical team involving all relevant specialists. Just as optimal oncologic practice currently brings together appropriate experts (medical oncologists, surgical oncologists, radiation oncologists, pathologists, radiologists and others as needed) in formal tumor boards, the genomics equivalent in the form of “molecular tumor boards” would improve medical management of patients. Specialists would include clinicians in relevant specialties, clinical genomicists, molecular biologists, bioinformaticists, and statisticians.

Moreover, guidelines must be formulated in order to provide uniform standards for genomics interpretations with regard to disease risk, genomically targeted interventions and pharmacogenomic considerations. The enforcement of uniform standards would provide community consensus, which would standardize the downstream interpretation of genomics findings. The increasing use and emphasis of electronic medical records (EMRs) in routine clinical care must incorporate genomic information (27). EMRs must include fields pertaining to genetic/genomic screening and interpretation, as for example family history, variants with comments on interpretation, and actionable findings. Approaches and procedures should be shared across institutions and groups within institutions (27). In sum, this will require the bringing together of what is now largely a fragmented array of genetic/genomic expertise under a single unifying institutional entity.

Establishment of an infrastructure that oversees and requires interaction among disciplines, in the fashion of TCGA and other large initiatives, would significantly benefit the implementation of genomics projects. Formal interdisciplinary programs offering cross-training in quantitative areas such as computational biology/genome science with biological and clinical science have been introduced and address many of the obstacles to communication. But such cross-disciplinary programs need to be established more broadly. At a minimum, researchers must be trained in multiple domains because these individuals will have the ability to communicate across areas of expertise. Over and above formal cross-disciplinary training programs, communication among the various specialists would benefit from the establishment of cme-accredited genomics training programs. Actual implementation of this approach is already evident in a variety of initiatives promoted by the National Cancer Institute, including those contained in the Cancer Moonshot Research Initiative (28). Although these recommendations are by no means comprehensive, they attempt to present the critical components of a new systematic approach to genomic medicine. Work which continues to be carried out in some institutions in isolated communities, essentially “siloes”, would be enriched by implementation in a prospectively structured collaborative environment. More immediate remedies to improvements in genomics literacy are critical as well. To sum up, these should include the integration of molecular tumor boards into the clinical programs, inclusion of clinical genomicists, and redesigning the EMRs to include relevant genomics fields, and improvement of content and accessibility of online resources.

References:

1. Cheng ML, Solit DB. Opportunities and Challenges in Genomic Sequencing for Precision Cancer Care. *Ann Intern Med.* 2018;168(3):221–2. [PubMed: 29310131]
2. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials. *Nature.* 2013;502(7471):317–20. [PubMed: 24132288]
3. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. *Stat Med.* 2014;33(11):1946–78. [PubMed: 24399688]
4. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol.* 2017;35(4):316–9. [PubMed: 28398311]
5. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: The perils of multiple testing. *Perspect Clin Res.* 2016;7(2):106–7. [PubMed: 27141478]
6. Faris J, Kolker E, Szalay A, Bradlow L, Deelman E, Feng W, et al. Communication and data-intensive science in the beginning of the 21st century. *Omics: a journal of integrative biology.* 2011;15(4):213–5. [PubMed: 21476843]
7. TCGA. TCGA/The Cancer Genome Atlas Program [updated March 6, 2019. Available from: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>.
8. ICGC. ICGC (International Cancer Genome Consortium) 2007 [updated June 14, 2018. Available from: <https://icgc.org/>.
9. APOLLO. APOLLO (Applied Proteogenomics Organizational Learning and Outcomes): NIH, National Cancer Institute, Office of cancer Clinical Proteomics Research, Center for Strategic Scientific Initiatives; 2016 [Available from: <https://proteomics.cancer.gov/programs/apollo-network>.
10. Bohannon ZS, Mitrofanova A. Calling Variants in the Clinic: Informed Variant Calling Decisions Based on Biological, Clinical, and Laboratory Variables. *Comput Struct Biotechnol J.* 2019;17:561–9. [PubMed: 31049166]
11. von Arnim AG, Missra A. Graduate Training at the Interface of Computational and Experimental Biology: An Outcome Report from a Partnership of Volunteers between a University and a National Laboratory. *CBE Life Sci Educ.* 2017;16(4).
12. Xu C A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput Struct Biotechnol J.* 2018;16:15–24. [PubMed: 29552334]
13. Hofmann AL, Behr J, Singer J, Kuipers J, Beisel C, Schraml P, et al. Detailed simulation of cancer exome sequencing data reveals differences and common limitations of variant callers. *BMC Bioinformatics.* 2017;18(1):8. [PubMed: 28049408]
14. Clark C, Kalita J. A comparison of algorithms for the pairwise alignment of biological networks. *Bioinformatics.* 2014;30(16):2351–9. [PubMed: 24794929]
15. Barnell EK, Ronning P, Campbell KM, Krysiak K, Ainscough BJ, Sheta LM, et al. Standard operating procedure for somatic variant refinement of sequencing data with paired tumor and normal samples. *Genet Med.* 2019;21(4):972–81. [PubMed: 30287923]
16. Richardson S, Tseng GC, Sun W. Statistical Methods in Integrative Genomics. *Annu Rev Stat Appl.* 2016;3:181–209. [PubMed: 27482531]
17. Eden C, Johnson KW, Gottesman O, Bottinger EP, Abul-Husn NS. Medical student preparedness for an era of personalized medicine: findings from one US medical school. *Per Med.* 2016;13(2):129–41. [PubMed: 27528879]
18. Krier JB, Kalia SS, Green RC. Genomic sequencing in clinical practice: applications, challenges, and opportunities. *Dialogues Clin Neurosci.* 2016;18(3):299–312. [PubMed: 27757064]
19. Salari K, Karczewski KJ, Hudgins L, Ormond KE. Evidence that personal genome testing enhances student learning in a course on genomics and personalized medicine. *PLoS One.* 2013;8(7):e68853. [PubMed: 23935898]
20. Su P Direct-to-consumer genetic testing: a comprehensive view. *Yale J Biol Med.* 2013;86(3):359–65. [PubMed: 24058310]

21. van der Wouden CH, Carere DA, Maitland-van der Zee AH, Ruffin MTt, Roberts JS, Green RC, et al. Consumer Perceptions of Interactions With Primary Care Providers After Direct-to-Consumer Personal Genomic Testing. *Ann Intern Med.* 2016;164(8):513–22. [PubMed: 26928821]
22. Kaphingst KA, McBride CM, Wade C, Alford SH, Reid R, Larson E, et al. Patients' understanding of and responses to multiplex genetic susceptibility test results. *Genet Med.* 2012;14(7):681–7. [PubMed: 22481132]
23. Feero WG. Introducing “Genomics and Precision Health”. *JAMA.* 2017;317(18):1842–3. [PubMed: 28492884]
24. NHGRI. National Human Genome Research Institute. for Health Professionals 2018 [updated December 29, 2018. Available from: <https://www.genome.gov/health/For-Health-Professionals>.
25. WHO. Human Genetics programme. Resources for health professionals World Health Organization Avenue Appia 20 CH-1211 Geneva 27 Switzerland [Available from: <https://www.who.int/genomics/professionals/en/>.
26. Manolio TA, Chisholm RL, Ozenberger B, Roden DM, Williams MS, Wilson R, et al. Implementing genomic medicine in the clinic: the future is here. *Genet Med.* 2013;15(4):258–67. [PubMed: 23306799]
27. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med.* 2010;363(2):166–76. [PubMed: 20647212]
28. Moonshot C National Cancer Institute. Cancer Moonshot Research Initiative. 21st Century Cures Act NCI website: National Cancer Institute; 2016 [Available from: <https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative>.

Figure 1: Unidirectional Approach to Transfer of Information among Experts in Relevant Disciplines during Implementation of a Genomic Pipeline: a Linear Approach

The traditional approach to information transfer as depicted in the figure reflects the siloed communication or non-communication among experts in required disciplines contributing to genomic analyses of biological material.

Sources of variability - symbols:

- ⊙ Technical variability:
 - ⊙ Use of fresh tissue *versus* FFPE (formalin fixed paraffin embedded) tissue; amount of input DNA
 - ⊙ Extent of DNA fragmentation
- ⊙ Bioinformatics variability-algorithm selection: (see ⊙ for abbreviation definitions)
 - ⊙ Algorithm selection for alignment to reference sequence
 - ⊙ Algorithm selection for genetic alteration identification

⊕ Clinical interpretation of DNA motifs: SNVs (single nucleotide variants), SNPs (single nucleotide polymorphisms), VUS (variants of uncertain/unknown significance), INDELs (insertion/deletions), CNVs (copy number variants), fusion variants

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2: Multidirectional Approach to Transfer of Information among Experts in Relevant Disciplines during Implementation of a Genomic Pipeline: a Network Approach.

Legend: In contrast to the linear approach (Figure 1), the multidirectional approach releases the constraints of a unidirectional pipeline. This breaking down of the siloes historically have isolated domain experts should enable improved communication and ultimately lead to better quality science.

Sources of variability - symbols: see Figure 1 for symbol definitions See Figure 1 Legend © for abbreviation definitions