Published in final edited form as:

Methods Mol Biol. 2019 ; 1908: 61–71. doi:10.1007/978-1-4939-9004-7_5.

Biological Interpretation of Complex Genomic Data

Kathleen M. Fisch

Department of Medicine, Center for Computational Biology and Bioinformatics, University of California San Diego, La Jolla, CA, USA.

Abstract

Tumor genomic profiling involves analyzing many data types to produce a molecular profile of a tumor. Many of these analyses result in a prioritized list of genes or variants for further study. Interpretation of these lists relies upon annotating and extracting biological meaning through literature and manually curated knowledge bases. This chapter will describe several of these approaches including gene annotation, variant annotation, clinical annotation, functional enrichment analyses, and network analyses. Taken together or individually, these analyses will result in a biological understanding of complex genomic data to improve clinical decision making.

Keywords

Computational biology; Bioinformatics; Variant annotation; Pathway analysis; Network analysis; Functional enrichment; Genomic interpretation

1 Introduction

The wealth of genomic data obtained from next generation sequencing technologies empowers researchers to create a molecular portrait of a patient's tumor [1–3]. However, after primary data analysis, data must be annotated, prioritized, and interpreted to be clinically relevant [4]. Various approaches can be implemented to reach this end, including gene and variant annotation, gene set enrichment analysis, pathway analysis, and network analysis [5].

Gene and variant annotations attach biologically curated knowledge to these entries to distill functional and mechanistic information [6–9]. Tools such as [MyGene.info](#) [8], BioGPS [7], ANNOVAR [6], and [MyVariant.info](#) [9] enable variant annotation by compiling information from various databases. Clinical databases exist to attach prognostic [10], drug [11, 12], and clinical trial information ([ClinicalTrials.gov](#)) to individual genes or variants. Canonical pathway and functional enrichment analyses enable the functional annotation of gene groups or variants, using tools such as ToppGene [13] and WebGestalt [14]. Finally, network analyses are used to examine interactions between genes to help predict the function of gene sets and to identify neighboring genes [15]. All of these analysis types heavily depend upon biological knowledge bases. Cancer specific knowledge bases exist, such as COSMIC [16]

and cBioPortal [10], which catalog known cancer associated variants and those found in The Cancer Genome Atlas (TCGA) [17].

Taken together, these tools allow for the creation of a molecular portrait that place genomic data into a biological context for clinical interpretation and actionability. In this chapter, three methods for biological interpretation of complex genomic data are described (Fig. 1). Subheading 2 describes the required hardware, software, and input data to implement these analyses. Subheading 3.1 describes annotating genes and variants with functional prediction algorithms and databases to evaluate the impact of a genomic alteration. Subheading 3.2 describes clinical annotation databases that can be used to identify relevant drugs, prognoses and clinical trials. Subheading 3.3 describes functional enrichment, pathway and network analyses that can be used to evaluate the biological relevance of gene sets, finding connections between genes and for visualizing the results. This chapter aims to serve as a guide for downstream biological interpretation of complex genomic data.

2 Materials

1. **Hardware.** A computer connected to the Internet is sufficient to explore the examples presented here.
2. **Software.** Examples of the described analyses are accessible from the Internet. Examples of these tools as part of computational workflow analysis notebooks are described in the Notes.
3. **Input Data.** The methods described herein depend upon a list of differentially expressed genes (HUGO gene symbols [18], Entrez [19], or Ensembl [20] identifiers) and/or a list of variants (HGVS identifiers [21] or a VCF file).

3 Methods

3.1 Annotation of Genes and Variants

3.1.1 Gene Annotation—Differential expression and genomic analyses generally produce a list of genes of interest for further evaluation. The first step to interpret these gene lists is to annotate the individual genes [9]. There are several different gene identifiers that exist [18–20]. Translating between the identifiers is a nontrivial task, as there are different annotation versions and synonyms that need to be taken into account [9]. Several tools exist for the purpose of translating gene identifiers [7, 8, 22]. [MyGene.info](#) is a Web service that allows the user to query genes and obtain up-to-date gene annotations [7, 8]. It also includes an application programming interface (API) that can be used to programmatically query a list of genes as part of a workflow. For an example of using [MyGene.info](#) as part of a computational workflow, refer to Note 1.

To manually use [MyGene.info](#) to translate a gene ID, perform the following steps:

1. Navigate through a Web browser to [MyGene.info](#).
2. Click on the “Try live API now” button.
3. Click on “[MyGene.info](#) gene query service.”

4. In the drop-down menu, click “Get Gene query service.”
5. Type the gene symbol “BRCA2” in the first box in the “q” parameter field.
6. Click “Try it!” at the bottom of this drop-down menu.
7. Scroll down to view the results under “Response Body” to obtain all of the information returned for this query gene. Returned values include the Entrez Gene ID, gene name, gene symbol and taxonomy id.

To manually use [MyGene.info](#) to provide annotation for a gene, perform the following steps:

1. Navigate through a Web browser to [MyGene.info](#).
2. Click on the “Try live API now” button.
3. Click on “[MyGene.info](#) gene annotation services.”
4. Type in the BRCA2 Entrez ID (675) into the “geneid” box.
5. Click “Try it!” at the bottom of this drop-down menu.
6. Scroll down to view the results under “Response Body” to obtain all of the information returned for this query gene.

[MyGene.info](#) powers BioGPS, a gene annotation portal that provides a visual interface for gene annotation [7]. To use BioGPS to query a gene or list of genes, perform the following steps:

1. Navigate through a Web browser to [biogps.org](#).
2. Enter a gene or list of genes in the “Search genes here” box or click “Gene Symbols” under “Example Searches.”
3. Click the “Search” button.
4. Click the record in the results table you want to view.
5. Explore the results for that record.

3.1.2 Variant Annotation—Variant calling pipelines result in a list of variants, usually in a Variant Calling Format (VCF) file. VCF files are text files that contain meta-information and positional information for each variant detected. Annotating variants with gene association, functional impact, population frequency, disease relevance, and other information is necessary for prioritizing variants and biological interpretation [5]. Many tools are available to annotate variants, including ANNOVAR [6] and [MyVariant.info](#) [9]. ANNOVAR functionally annotates genetic variants with gene-based annotation to identify protein-coding changes and affected amino acids, region-based annotation such as conserved genomic regions among species, and predicted transcription factor binding sites, among others, and filter-based annotation that includes data from variant databases, population allele frequencies from whole genome and whole exome datasets, and functional predictions scores from a variety of sources [6]. Here we will demonstrate how to use [MyVariant.info](#). [MyVariant.info](#) is a variant annotation service that curates variant information from 20 sources to date and keeps them regularly updated [9]. Examples of the databases

[MyVariant.info](#) curates include dbNSFP [23], dbSNP [24], ClinVar [25], CADD [26], and COSMIC [16], among others. It also provides an API for programmatic annotation of variants. For an example of using [MyVariant.info](#) as part of a computational variant annotation and filtering workflow using the Python tool VAPr (<https://github.com/ucsd-cbb/VAPr>), refer to Note 2 [39]. To manually use [MyVariant.info](#) to query a variant, perform the following steps.

1. Create HGVS ID [21] from variant (Example: chr7: 5241707G>T).
2. Navigate to [MyVariant.info](#).
3. Click on the “Try live API now” button.
4. Click on “[MyVariant.info](#) variant annotation services.”
5. Click on “Get Variant annotation service.”
6. Input HGVS ID from step 1 in the “variantid” text field.
7. Click “Try it!” button at the bottom.
8. Scroll down to view the results under “Response Body,” returned as a JSON document, to obtain all of the information returned for this query variant.

3.2 Clinical Annotation

3.2.1 Targeted Therapeutic Databases—Tumor profiling often has the goal of identifying clinically actionable genes or variants that can be targeted therapeutically [27]. Several resources exist that contain drug information. One example is the drug-gene interaction database (DGIdb) [11, 12], which curates drug-gene interactions from several sources such as DrugBank [28], PharmGKB [29], and [ClinicalTrials.gov](#) using a combination of expert curation and text-mining. DGIdb requires a list of genes and will return all druggable or potentially druggable genes from that list. Perform the following steps to explore the functionality of DGIdb:

1. Navigate to DGIdb (<http://dgidb.genome.wustl.edu/>) through a Web browser.
2. Select the “Search Drug-Gene Interactions” button.
3. Enter one or more genes in the textbox, or click the “Replace with demo list” button at the bottom of the textbox.
4. Click the check boxes under “Preset Filters” and toggle the menus under “Advanced Filters” (“Source Databases,” “Gene Categories,” and “Interaction Types”) to view the available options for each category.
5. Click the button “Find Drug-Gene Interactions” at the bottom of the screen.
6. Explore the Results Summary of Drug-Gene interactions.

3.2.2 Clinical Trials—In addition to identifying druggable targets from a tumor molecular profile, clinical trials relevant to the observed genomic alterations can be searched. [ClinicalTrials.gov](#) is a Web-based database maintained by the National Library of Medicine and the National Institutes of Health that contains information on publicly and

privately supported clinical trials. Study protocol information for each clinical trial includes the disease or condition under study, type of intervention being studied, the title, description and design of the study, participant eligibility criteria, study location, and contact information. Other relevant information includes description of study participants, study outcomes and experienced adverse events. To search [ClinicalTrials.gov](https://clinicaltrials.gov) for clinical trials relevant to a list of genes or mutations from a tumor profile, perform the following steps:

1. Navigate to <https://clinicaltrials.gov/> through a Web browser.
2. Locate the textbox under “Find a study.”
3. Input a gene name, such as “BRCA2” into the “Other terms” textbox and click “Search.”
4. Explore all of the relevant clinical trial results.
5. Click on the name of the clinical trial for additional information.

3.2.3 Prognostic Annotation—The large amount of cancer patient data curated by projects such as The Cancer Genome Atlas (TCGA) enables researchers to analyze clinical and genomic data for 33 cancer types from more than 11,000 patients [17]. Data types include gene expression, somatic mutations, DNA methylation, copy number variation, protein expression and clinical information. Many groundbreaking studies have been done on the TCGA dataset [2, 17, 30, 31], which have contributed to our cancer knowledge base. Tools such as cBioPortal [10] provide easy to use Web interfaces to visually explore analyzed TCGA data. cBioPortal allows a user to search by specific cancer type and by genes of interest. It returns cohort statistics and prognostic annotation, including survival analysis based on alterations in genes of interest. To explore cBioPortal for a list of genes of interest, perform the following steps:

1. Navigate to <http://www.cbioportal.org/> through a Web browser.
2. Locate the “Query” tab.
3. Search for a cancer study or select a check box for the cancer study of interest.
4. Click on the check boxes next to data types that you are interested in viewing in the “Select Genomic Profiles” box.
5. From the drop-down menu in the “Select Patients/Case Set” select the dataset of interest.
6. In the “Enter Gene Set” box, type in HUGO gene symbols that you are interested in or select a precompiled gene list from the drop-down menu to select genes to view.
7. Click the “Submit Query” button.
8. Explore the summary results in the “Overview” tab summarizing the alterations in the genes of interest.
9. Click on the “Mutations” tab to view details about the mutations in the genes of interest.

10. Click on the “Expression” tab to view expression levels for the genes of interest.
11. If you choose a single cancer type from the Query tab, you will be able to view detailed analyses for the genes of interest located in additional tabs, including “Mutual Exclusivity,” “Co-Expression,” “Enrichments,” “Survival,” and “Network.”

3.3 Functional Enrichment Analysis

3.3.1 Gene Set Enrichment Analysis—Annotating individual genes and variants as described above is complemented by a gene set enrichment approach, which detects functional enrichment of a gene list based on biologically curated gene sets. This allows for biological interpretation of complex data sets and tumor profiles by identifying significantly enriched biological entities (Gene Ontology terms, canonical pathways, drug targets, etc.) for the gene set of interest. Many tools exist for performing gene set enrichment analysis, including the Web-based tool ToppFun, part of the ToppGene Suite of tools, which detects functional enrichment of a gene list for 18 biologically curated and up-to-date gene sets [13]. For an example of using the ToppFun API as part of a computational workflow, refer to Note 3. To perform a functional enrichment analysis using ToppFun, perform the following steps:

1. Navigate to <https://toppgene.cchmc.org/> through a Web browser.
2. Click on “ToppFun.”
3. Enter a list of HUGO gene symbols or click on the “Example gene sets” link.
4. Click “Submit.”
5. Review the list of input genes and revise any that are not found at the top under “Input Gene List.”
6. View the “Feature” gene sets that will be used in the analysis under the “Calculations” header. You can leave all of the boxes checked and with default values.
7. Click “Start” at the bottom of the page to run the analysis.
8. View the results on the “Results” page. Results are divided into gene set source and are ranked by p-values. You can download a spreadsheet of the results by clicking on the “Download All” link at the top. You can view the “Genes from Input” and “Genes in Annotation” interactively from the Web page by clicking on those links in the results.

3.3.2 Pathway Analysis and Visualization—Pathway analysis is a gene set enrichment analysis using canonical pathways as the gene set to test for significant enrichment [32]. ToppFun includes pathway databases and returns pathway analysis results [13]. Another useful tool for performing pathway analyses and visualizing the input genes overlaid onto the canonical pathway map is WebGestalt [14, 33]. For an example of visualizing dysregulated pathways as part of a computational workflow, refer to Note 3. To perform a KEGG pathway analysis in WebGestalt to view the input genes overlaid on the pathway map, perform the following steps:

1. Navigate to www.webgestalt.org through a Web browser.
2. Select the “organism of interest” from the drop-down list. For this example, choose “hsapiens.”
3. Select the “method of interest” from the drop-down list. For this example, choose “Overrepresentation Enrichment Analysis.”
4. Select the “functional database” from the drop-down list. For this example, choose “pathway” under “functional database class” and “KEGG” under “functional database name.”
5. From the drop-down menu under “Select gene ID type” choose “genesymbol.”
6. Enter your list of gene symbols in the text box under “Upload gene list.”
7. Under the “Select Reference Set for Enrichment Analysis” choose “genome_protein-coding” to use all protein coding genes as the background for the Hypergeometric test. For a real experiment, it is advised to upload a reference background of genes expressed/tested in your study.
8. Leave the “Advanced Parameters” at the default levels or adjust as desired.
9. Click “Submit.”
10. Click on “View Results.”
11. Explore the results table listing the significantly enriched pathways.
12. To view the input genes overlaid on the canonical pathway, click the name of the pathway. The genes in this pathway will appear on the right side of the screen. Click on the pathway name above the genes and you will be taken to the KEGG website. Genes from the input query will be colored red in the resulting diagram.

3.3.3 Network Analysis—Network analyses are used to examine interactions between genes to help predict the function of gene sets and to identify neighboring genes [15]. Several tools exist to perform these analyses, such as Cytoscape [34], String [35, 36], and GeneMania [37]. This example network analysis will be demonstrated using a Web-based tool, GeneMania [37]. GeneMania takes a list of genes as input and finds other genes that are related to this set through a large set of functional association data, such as protein and genetic interactions, canonical pathways, coexpression, colocalization, and protein domain similarity [37]. It also allows you to perform functional enrichment analyses for Gene Ontology (GO) [38] terms on the genes in your query GeneMania network. For advanced network analysis methods, refer to Note 4. To perform a network analysis with GeneMania, perform the following steps.

1. Navigate to genemania.org through a Web browser.
2. Identify the black bar at the top left of the screen and click within the white input textbox.
3. Enter a list of gene symbols, one per line.
4. Click the search icon to the right of the textbox.

5. Explore the network result by interacting with the network diagram. You can hover over the gene nodes or drag them around.
6. View the edge types and sources in the “Networks” menu to the right of the screen. Toggle edge types on and off by clicking the checkbox.
7. Click the icons to the left of the screen to create different network views, export the network to a file and get more information about the network.
8. View the functional enrichment analysis of your network by clicking the icon in the lower left corner. Click the check boxes next to GO terms of interest to highlight network genes in this GO term.

4 Notes

1. To view code that includes a gene id conversion step with [MyGene.info](https://mygene.info), please refer to the Jupyter-Genomics GSEAPy notebook: <https://github.com/ucsd-ccbb/jupyter-genomics/tree/master/notebooks/rnaSeq>.
2. To view code for a full variant annotation and filtering workflow, please refer to the VAPr Variant Annotation and Prioritization Github repository: <https://github.com/ucsd-ccbb/VAPr> [39].
3. To view code for programmatic implementation of functional enrichment analyses, please refer to the Jupyter-Genomics Functional Enrichment Analysis and Pathway Visualization notebook: <https://github.com/ucsd-ccbb/jupyter-genomics/tree/master/notebooks/rnaSeq>.
4. To view code for advanced network analyses, please refer to the visJS2jupyter Github repository: <https://github.com/ucsd-ccbb/visJS2jupyter> [40].

References

1. Hoadley KA, Yau C, Wolf DM et al. (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158:929–944 [PubMed: 25109877]
2. Kandoth C, McLellan MD, Vandin F et al. (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502:333–339 [PubMed: 24132290]
3. Tamborero D, Gonzalez-Perez A, Perez-Llamas C et al. (2013) Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 3:2650 [PubMed: 24084849]
4. Dienstmann R, Dong F, Borger D et al. (2014) Standardized decision support in next generation sequencing reports of somatic cancer variants. *Mol Oncol* 8:859–873 [PubMed: 24768039]
5. Moreau Y, Tranchevent LC (2012) Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat Rev Genet* 13:523–536 [PubMed: 22751426]
6. Wang K, Li M, Hakonarson H (2010) ANNO-VAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res* 38(16):e164 [PubMed: 20601685]
7. Wu C, MacLeod I, Su A (2013) BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res* 41: D561–D565 [PubMed: 23175613]
8. Wu CW, Mark A, Su AI (2014) MyGene.Info: gene annotation query as a service. *bioRxiv*. 10.1101/009332
9. Xin J, Mark A, Afrasiabi C et al. (2016) Highperformance web services for querying gene and variant annotation. *Genome Biol* 17:1–7 [PubMed: 26753840]

10. Gao J, Aksoy BA, Dogrusoz U et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269):p11 [PubMed: 23550210]
11. Griffith M, Griffith OL, Coffman AC et al. (2013) DGIdb: mining the druggable genome. *Nat Methods* 10:1209–1210 [PubMed: 24122041]
12. Wagner AH, Coffman AC, Ainscough BJ et al. (2016) DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res* 44: D1036–D1044 [PubMed: 26531824]
13. Chen J, Bardes E, Aronow B et al. (2009) ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37:W305–W311 [PubMed: 19465376]
14. Wang J, Duncan D, Shi Z et al. (2013) WEB-based GENE SeT analysis toolkit (Web- Gestalt): update 2013. *Nucleic Acids Res* 41: W77–W83 [PubMed: 23703215]
15. Mitra K, Carvunis AR, Ramesh SK et al. (2013) Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* 14:719–732 [PubMed: 24045689]
16. Forbes SA, Bindal N, Bamford S et al. (2011) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 39:D945–D950 [PubMed: 20952405]
17. Weinstein JN, Collisson EA, Cancer Genome Atlas Research Network et al. (2013) The cancer genome atlas pan-cancer analysis project. *Nat Genet* 45:1113–1120 [PubMed: 24071849]
18. Eyre TA, Ducluzeau F, Sneddon TP et al. (2006) The HUGO gene nomenclature database, 2006 updates. *Nucleic Acids Res* 34: D319–D321 [PubMed: 16381876]
19. Brown GR, Hem V, Katz KS et al. (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res* 43:D36–D42 [PubMed: 25355515]
20. Flicek P, Ahmed I, Amode MR et al. (2013) Ensembl 2013. *Nucleic Acids Res* 41: D48–D55 [PubMed: 23203987]
21. den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15:7–12 [PubMed: 10612815]
22. Smedley D, Haider S, Durinck S et al. (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43(W1): W589–W598 [PubMed: 25897122]
23. Liu X, Jian X, Boerwinkle E (2013) dbNSFP v2. 0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 34:E2393–E2402 [PubMed: 23843252]
24. Sherry ST, Ward MH, Kholodov M et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311 [PubMed: 11125122]
25. Landrum MJ, Lee JM, Riley GR et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42:D980–D985 [PubMed: 24234437]
26. Kircher M, Witten DM, Jain P et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46:310–315 [PubMed: 24487276]
27. Van Allen EM, Wagle N, Stojanov P et al. (2014) Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffinembedded tumor samples to guide precision cancer medicine. *Nat Med* 20:682–688 [PubMed: 24836576]
28. Law V, Knox C, Djoumbou Y et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 42: D1091–D1097 [PubMed: 24203711]
29. Hewett M, Oliver DE, Rubin DL et al. (2002) PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res* 30:163–165 [PubMed: 11752281]
30. Ciriello G, Miller ML, Aksoy BA et al. (2013) Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 45:1127–1133 [PubMed: 24071851]
31. Guo Y, Sheng Q, Li J et al. (2013) Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS One* 8:e71462 [PubMed: 23977046]
32. Ramanan VK, Shen L, Moore JH et al. (2012) Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* 28:323–332 [PubMed: 22480918]
33. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33:W741–W748 [PubMed: 15980575]

34. Cline MS, Smoot M, Cerami E et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2:2366–2382 [PubMed: 17947979]
35. Franceschini A, Szklarczyk D, Frankild S et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41:D808–D815 [PubMed: 23203871]
36. Snel B, Lehmann G, Bork P et al. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* 28:3442–3444 [PubMed: 10982861]
37. Zuberi K, Franz M, Rodriguez H et al. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res* 41:W115–W122 [PubMed: 23794635]
38. Blake JA, Dolan M, Gene Ontology Consortium et al. (2013) Gene ontology annotations and resources. *Nucleic Acids Res* 41: D530–D535 [PubMed: 23161678]
39. Birmingham A, Mark AM, Mazzaferro C, Xu G, Fisch KM (2018) Efficient population-scale variant analysis and prioritization with VAPr. *Bioinformatics* 34(16):2843–2845 [PubMed: 29659724]
40. Rosenthal SB, Len J, Webster M, Gary A, Birmingham A, Fisch KM (2018) Interactive network visualization in Jupyter notebooks: visJS2jupyter. *Bioinformatics* 34(1):126–128 [PubMed: 28968701]

Fig. 1.
Biological interpretation analysis landscape for complex data