Land Explorer for QIAGEN® Ingenuity® Pathway Analysis (IPA®)

Connect the dots: Deriving insights from thousands of ‘omics datasets

Sample to Insight

With the rise of 'omics technologies and large-scale consortia projects, biological systems are studied at an unprecedented scale, generating enormous numbers of heterogeneous and often large datasets. These high-throughput technologies enable 'omics studies that evaluate thousands of genes and millions of variants. Approaches to combine many thousands of datasets spanning multiple 'omics types provide a more comprehensive understanding of complex genotype-phenotype associations than the analysis of a single dataset alone. To make this easier to do, Land Explorer for QIAGEN IPA brings together the data for over half a million 'omics samples into one intuitive interface.

It can be burdensome to compile, store, search, analyze and visualize all relevant information from massive repositories of 'omics experiments for medical conditions, such as how gene regulation changes by tissue, disease or treatment condition. Mining data and the literature can identify some key links; still, data relevant to research is often buried or hard to find, limiting the identification of meaningful connections. Researchers often end up 'cherry picking' the studies they use for comparisons, leading to results that may be biased in favor or against a particular association.

Yet, to explore these highly valuable but complex datasets constitutes not only a conceptual challenge but a practical hurdle in the daily analysis of 'omics data. Even with sufficient resources to download and store the data, it takes a combination of biology expertise to properly annotate various features related to the samples or subjects from which the data was collected, bioinformatics skills for reanalyzing datasets to be consistent with one another, and further resources to publish them to researchers in a comprehensible interface.

"The quintessential inquiry in the genomics era," postulates Dr. Ben Darbro, M.D., Ph.D., Associate Professor and Director of the Shivanand R. Patil Cytogenetics and Molecular Laboratory at the University of Iowa, "is, 'Here's my data, give me

"I use IPA quite a bit because there's nothing that has more information. There's nothing out there that has this kind of knowledge base behind it."

Dr. Ben Darbro, M.D., Ph.D., Associate Professor and Director of the Shivanand R. Patil Cytogenetics and Molecular Laboratory at the University of Iowa

an answer'. That's always been the question ever since we started to develop 'omics techniques. One gene, one hypothesis, one system – that's pretty easy to interpret. But anytime you get into the 'omics space—whether proteomics, transcriptomics, genomics—it's really hard to focus on just one thing in an analysis. Not to mention, you wouldn't want to do that, because you would be wasting all the rest of this information."

Ultimately, it is difficult and time-consuming to analyze, interpret and contextualize findings from 'omics experiments, which slows down scientific discovery. Therefore, there is a need for robust and advanced analysis strategies to harness the value of these comprehensive high-throughput data, to make it easier to identify true associations and reduce the number of false associations.

Land Explorer for QIAGEN IPA: A massive collection of 'omics data at your fingertips

Land Explorer for QIAGEN Ingenuity Pathway Analysis (IPA) is QIAGEN's web-based platform that helps life science researchers perform insightful data analysis and interpretation. It enables the understanding of experimental results within the context of various biological systems, and empowers contextualizing findings from QIAGEN IPA against more data than could ever be analyzed by a single person. This powerful tool dramatically accelerates data interpretation, hypothesis generation and, ultimately, scientific publication.

Land Explorer for IPA enables you to access a vast amount of information to identify gene signatures and explore 'omics data for individual genes and to visualize expression correlation across multiple genes. You can jump from a gene of interest identified in an analysis in QIAGEN IPA to discover its tissue or cell expression in other samples or datasets. For biologists that would like to draw on the vast number of 'omics data types, Land Explorer is a great resource that also serves as a central location to visualize these data in various ways, and works through web browsers, allowing the exploration and visualization of large amounts of content in one place (Table 1), without having to install separate software platforms. Not only does Land Explorer free up time and resources to create impactful visualizations with public datasets, but it also helps you develop and validate research hypotheses and provides evidence that can be used for publication.

> "The number one thing that stands out in QIAGEN IPA versus any other solution is that it's intuitive and easy to use."

To provide comprehensive systems biology analysis, Land Explorer facilitates the interrogation of findings in real time against a library of over 500,000 samples across tissues, treatments and diseases from thousands of 'omics experiments to identify unique or common biological features. You can explore a wide variety of 'omics datasets, including RNA-seq gene/transcript expression, RNA-seq gene fusion expression, expression microarrays, methylation microarrays, miRNA-seq, reverse phase protein lysate microarray (RPPA), somatic mutations, copy number alterations and curated metadata/survival data (Table 2). Land Explorer lets you survey interactive plots of gene expression in 51 human tissues for gene-level and individual splice variants and visualize patterns of differential transcript expression across multiple parameters (Figure 1), including cell type, cell line, tissue, disease, treatment, sample type and project or clinical metadata. Besides viewing a single data type in Land Explorer, correlation views let you integrate different data types for samples with, for example, both gene expression data and copy number variation data. With this solution for 'omics data analysis, you can inspect expression, mutation status, copy number and other details from any gene of interest using thousands of statistical comparisons from a consistent data analysis, processing, curation and statistical pipeline (Figure 2) to make true apples-to-apples comparisons between different studies. Such a feat is simply not possible using other data repositories.

Table 1. Data sources for Land Explorer*

Land Name	Land Type	Sample Count	No. Data Types
GTEx	BodyMap	17783	7
Blueprint	BodyMap	629	6
HPA	BodyMap	373	8
HumanRNAi	Cell line	1298	1
CellLine_GSK	Cell line	1289	2
CCLE	Cell line	1054	14
CellLine_Pfizer	Cell line	399	2
HumanCRISPR	Cell line	342	1
CellLine_NCI	Cell line	299	3
LINCS	Cell line	115210	1
HumanDisease	Disease	158996	10
MouseDisease	Disease	52422	9
RatDisease	Disease	9667	8
TCGA	Oncology	22687	14
ICGC	Oncology	11805	10
METABRIC	Oncology	4129	2
TARGET	Oncology	3942	12
expO	Oncology	2159	1
BeatAML	Oncology	666	8
CGCI	Oncology	110	6
OncoGEO	Oncology	63816	11
Hematology	Oncology	29492	12
TumorMutations	Oncology	21508	4
ClinicalOutcome	Oncology	11405	1
Pediatrics	Oncology	11030	10
MetastaticCancer	Oncology	4091	3
OncoMouse	Oncology	1733	8

*As of September 21, 2020

Table 2. Data types (and respective sample counts) available in Land Explorer

- RNA-seq gene/transcript expression, exons, junctions, fusions, mutations (70.000+)
- Expression microarrays (390.000+)
- Methylation microarrays (26.000+)
- miRNA-seq (16.000+)
- Somatic Mutations, copy number alterations (121.000+)
- Curated metadata/survival data

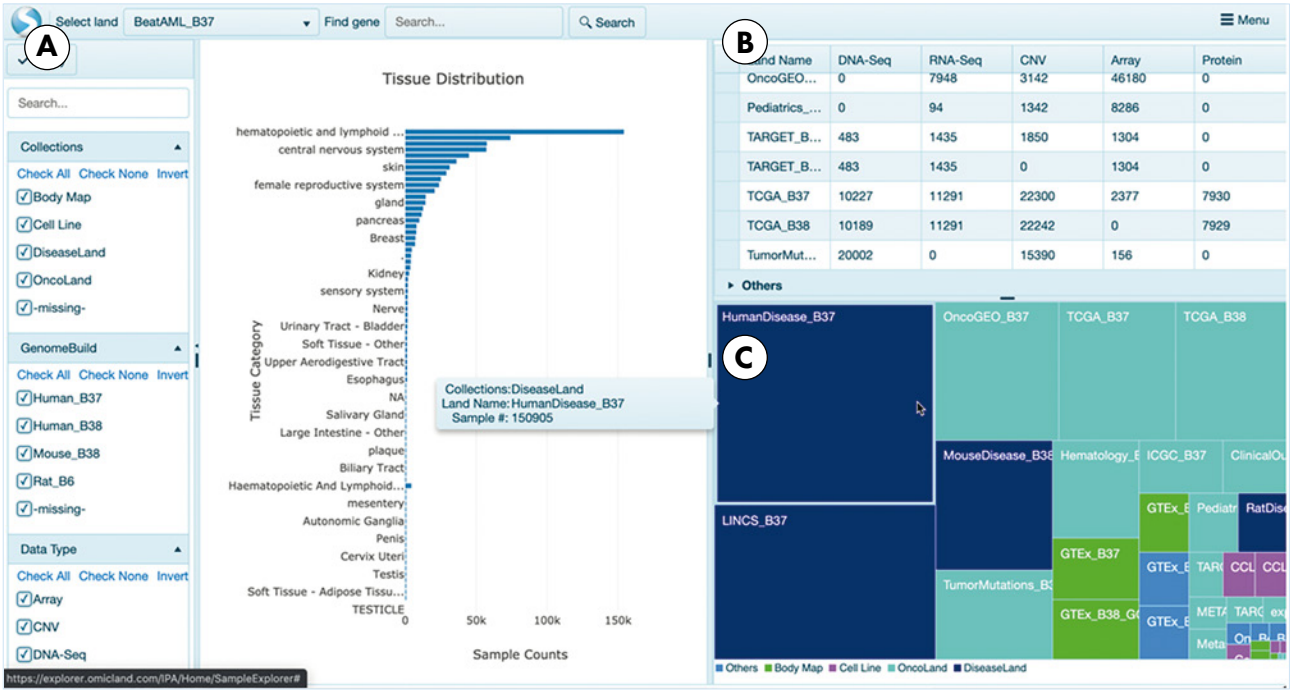


Figure 1. Overall sample distribution from Land Explorer. The Sample Explorer lets you browse data across all available lands based on collections or individual lands for samples or tissues of interest and combine all oncology- and disease-focused studies in one simple interface. The Sample Explorer lets you survey 548,000 'omics samples from OncoLand and DiseaseLand, including 15 consortium databases (TCGA, LINCS, GTEx, CCLE, ICGC, TARGET and more) containing 184,000 samples, 6000 independent projects, over 331,000 samples from individual publications, 33,000 samples in specialty Lands and over 60,000 statistical comparisons.

(A) Tissue distribution in Land Explorer. The tissue with the highest number of samples is blood, with over 150,000 samples and is followed by the central nervous system, which has 75,000 samples. **(B)** Land Explorer has multiple different experimental dataset types for a given sample set, including DNA-seq, RNA-seq, CNV, expression microarray and protein. **(C)** Tree block view of different sample types. The tree block view lets you filter down to samples or tissues of interest, identify which Land the samples are present within, and then quickly go into that Land that has already been pre-filtered to have the specific tissues that you would like to further investigate. **Green**: body map samples; **turquoise**: oncology-related samples; **dark blue**: non-cancer related disease samples; **purple**: cell line samples; **light blue**: non-categorized (other) samples.

Case study: Identify genetic determinants of disease

Dr. Darbro's research program focuses on the genetic determinants of disease and is split into two research groups: germline disorders and cancer. "I'm really interested in that intersection [between germline disorders and cancer], and I think one of the main areas of intersection is at a pathway level," Dr. Darbro says. "You think about things like conditions that lead to autism spectrum disorder or tuberous sclerosis or other germline disorders, but they're impacting pathways that are also critical for tumor suppression." Tools in IPA, such as Land Explorer, help reveal the intersection between these two worlds, and inform one another. Dr. Darbro says, "I think that's probably the real strength of looking at both germline disorders and cancer at the same time—that some of the pathways and tools that you become quite competent with, or developed, for one group can also be used to study the other group."

With this framework, Dr. Darbro's team also performs gene discovery in neurodevelopmental disorders like intellectual

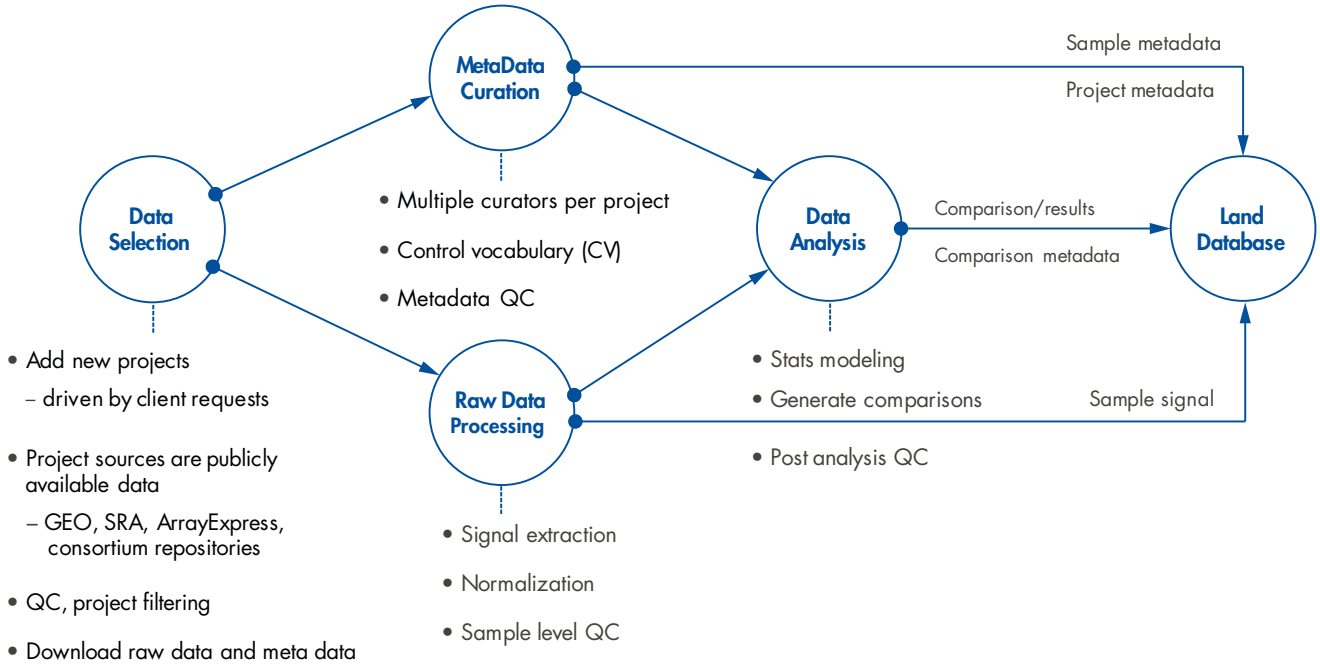


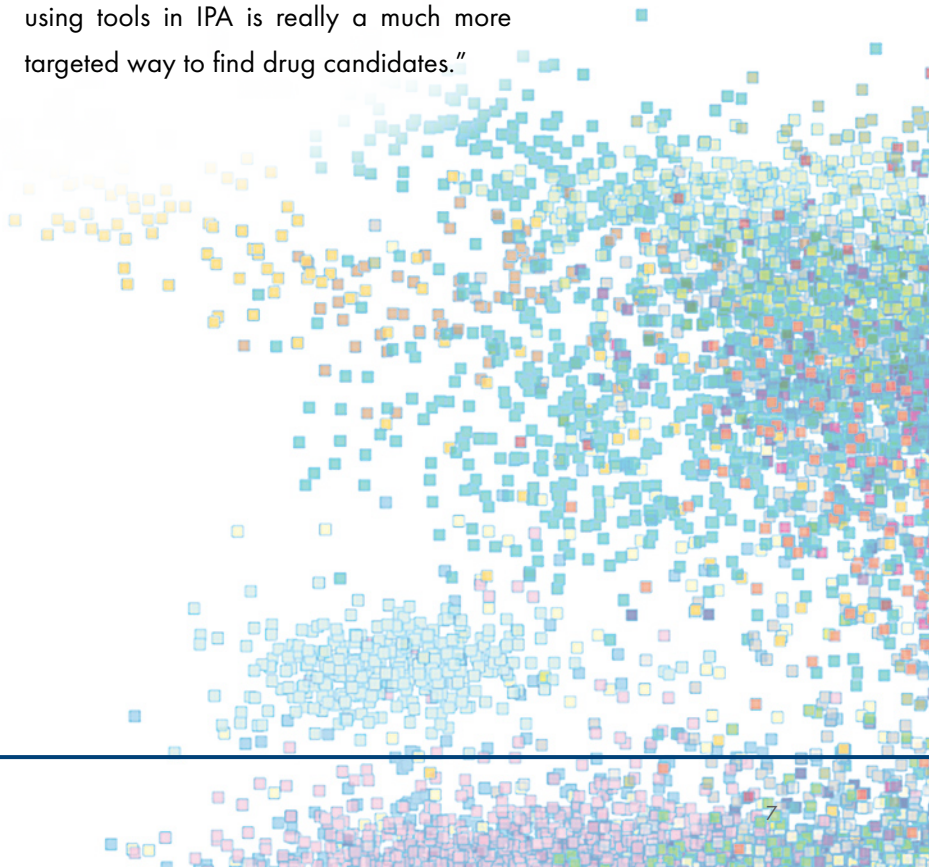
Figure 2. Land Data processing workflow. The 'omics datasets in Land Explorer are curated and re-processed, adding extra value to each dataset. The QIAGEN team invested over 42,000 person-hours in expert curation to curate the more than 500,000 samples contained in Land Explorer. Furthermore, it has taken an estimated 835,000 CPU-hours (or 95.3 CPU-years) from SRA download to full analysis, to analyze the 83,500 RNA-seq samples contained in Land Explorer. Finally, performing statistical analyses on the 325,000 samples used in comparisons converts to 9,750 person-hours of statistical analysis.

disability, autism spectrum disorder and epilepsy. "We're able to take a list of genes and start to try to see if there's a relationship between them. One of the things we did after running an IPA analysis on genes was use Land Explorer to try to figure out a little bit more about some of the upstream regulators and the canonical pathways that we found to be enriched. We try to branch out and say, 'all right, well, what other studies have seen this as well? What other studies have these pathways or these upstream regulators enriched?'"

To take the exploration of these genes and the underlying biology to the next level, Dr. Darbro's research groups look at the functional relevance of these genes and associated networks and identify biomarkers. "Not only is there an aspect of exploring those genes and how they directly interact with one another, but also understanding what functional networks they're a part of. If we have expression data and copy number data, we can use IPA's statistical framework to basically say that this particular copy number variant is functional because all of the genes in it change their expression in the right direction. "Dr. Darbro says these CNVs are often found in both germline diseases and cancer. "They're the kind of things that make for decent biomarkers because they're probably actually involved in the pathogenesis...as opposed to the myriad of 'bystander lesions' that show up."

In addition to discovering how genes are regulated in disease and respond to gene mutations, Dr. Darbro's laboratory uses tools in QIAGEN IPA, such as Land Explorer and Analysis Match, to go one step further to explore drug treatments. "The other way we use IPA—that I think is probably how we're going to continue to use it heavily—is in the drug selection area. There's no shortage of drugs out there. There are entire screening facilities that exist where you send them some cells and they test a thousand drugs to see what they do. But using tools in IPA is really a much more targeted way to find drug candidates."

> "Anytime you get into the 'omics space—whether proteomics, transcriptomics, genomics—it's really hard to focus on just one thing in an analysis. Not to mention, you wouldn't want to do that, because you would be wasting all the rest of this information."



How Land Explorer for IPA can help advance your research

Explore differential expression in human disease studies

Land Explorer for QIAGEN IPA tremendously accelerates how you examine gene expression in studies of human diseases to identify gene signatures and discover biomarkers. For instance, the gene *HNF1A* is a transcription factor highly expressed in the liver and regulates the expression of several liver-specific genes. In Land Explorer, there are many different comparisons where *HNF1A* is up- or down-regulated. Filters enable you to focus on specific human diseases, such as liver diseases (Figure 3). This reveals several studies that found significant differential regulation of *HNF1A*, such as those in hepatitis C, hepatitis B and acute liver failure. You can further survey these studies in more detail, such as all the genes differentially regulated in those studies. This ability to start with a specific gene of interest, then focus on a specific tissue,

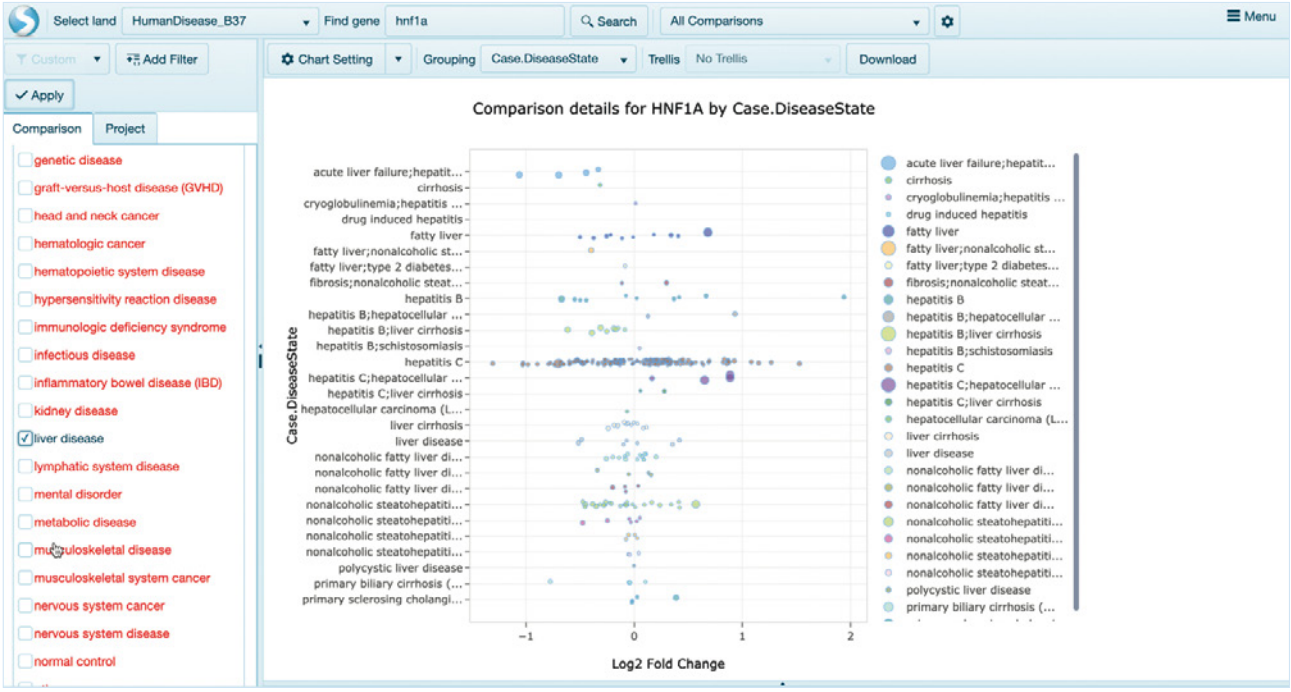


Figure 3. Comparison of the *HNF1A* gene expression by disease state. Statistical comparisons for differential expression of *HNF1A* in liver diseases using a bubble plot. *HNF1A* was found to be down-regulated compared to the control samples in acute liver failure, but different studies found *HNF1A* to be either up-regulated or down-regulated in hepatitis C, depending on the experimental conditions. X-axis represents differential gene expression of *HNF1A* by Log2 fold change. Y-axis represents individual disease states. Bubble size represents the significance (p-value) from the differential expression (DESeq2 for RNA-seq and General Linear Model for microarray) of *HNF1A*. Larger bubbles represent greater significance. Selecting dot(s) in this view will populate a Details table at the bottom of the view showing some key metadata from the comparison(s).

and then to include other significant genes, allows you to expand gene signatures to a whole set of genes in minutes. These genes can then be explored further in QIAGEN IPA to understand the biological underpinnings of the studies.

Explore gene expression in normal and disease cell types

When you want to assess 'omics data, such as gene expression, from different cell types, major difficulties include finding the right datasets to analyze, and processing these datasets to make them comparable. This can be resource-intensive, both in terms of funds and time, and becomes more and more difficult with the number of datasets that you include, since each one must be processed to make it comparable. Land Explorer for QIAGEN IPA helps you examine gene expression in normal and disease samples of specific cell types more efficiently. For example, you can explore gene expression in immune cells in both normal and tumor samples. Land Explorer allows you to visualize differences in expression across thousands of different samples with RNA-seq data, and filter down to the tissues (for example, blood, esophagus and nerves) or by cell types (normal or tumor samples of different immune cell types).

Discover survival rates for cancer gene mutations

Land Explorer for IPA enables the exploration of mutations for a particular gene, to get answers about how these mutations affect the expression of other genes, what diseases they may influence and how they might affect patient survival. For example, the tumor suppressor gene *PTEN* is found in most tissues, and mutations of this gene are involved in the development of many cancers and other diseases (Figure 4A). Land Explorer incorporates several public datasets to facilitate 'omics analysis and improve the level of generated insights. By featuring a collection of highly curated 'omics data, Land Explorer streamlines the analysis process to quickly focus efforts toward candidates with clear biological relevance to help you identify and explore contextual data in seconds. Using Land Explorer, you can survey the mutation prevalence across all the cancers within The Cancer Genome Atlas (TCGA), then subset samples from a particular cancer with a *PTEN* mutation from samples that do not. From here, Land Explorer can generate the survival plots for *PTEN* in samples that are wild-type versus cancer of interest (Figure 4B).

> "One of the things we did after running an IPA analysis on genes was use Land Explorer to try to figure out a little bit more about some of the upstream regulators and the canonical pathways that we found to be enriched. We try to branch out and say, 'what other studies have seen this as well?'"

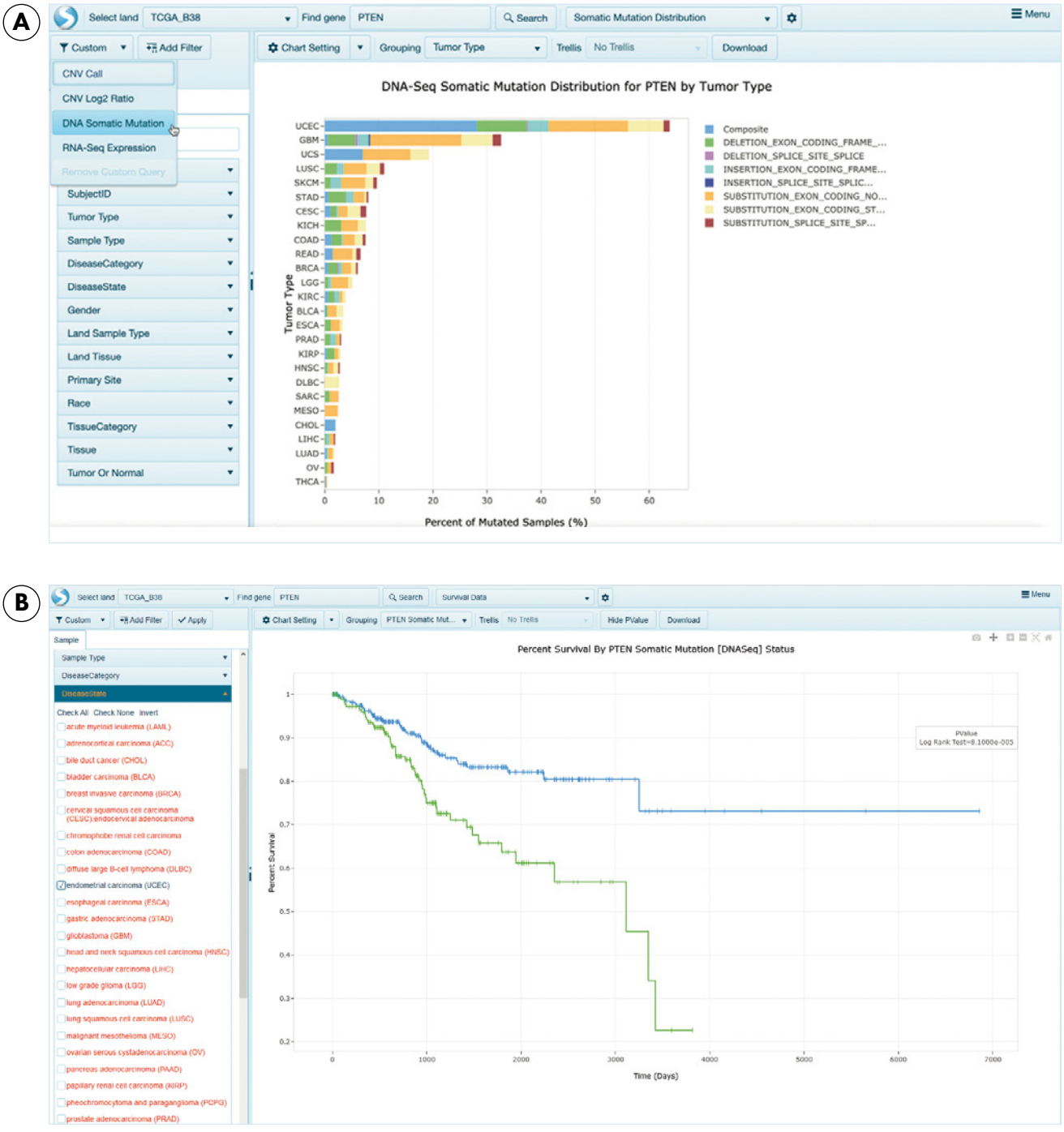


Figure 4. (A) DNA-seq somatic mutation distribution for *PTEN* by tumor type. Somatic mutation distribution of *PTEN* in tumor samples from the Cancer Genome Atlas (TCGA), including insertions, deletions and substitutions. PTEN is somatically mutated in over 60% of uterine corpus endometrial carcinoma (UCEC) samples and over 30% of glioblastoma multiforme (GBM) samples. X-axis represents the percent of mutated samples. **(B)** Survival plots for *PTEN* wild-type versus PTEN-mutation in cancer samples of interest. Survival and disease-free survival time of patients with or without specific gene mutations for *PTEN* were analyzed using Kaplan-Meier curves.

Explore genes implicated in a disease

Upon identifying a list of genes from an analysis in QIAGEN IPA, you can learn more about those genes individually in Land Explorer. For example, Land Explorer can aid in the study of the *ACE2* gene, which is required for the SARS-CoV-2 virus to enter cells and is responsible for the COVID-19 disease. Land Explorer for IPA contains a vast amount of valuable information, such as gene expression in normal tissue and disease tissue, or whether the gene is differentially over- or under-expressed across thousands of studies that have been curated and processed by QIAGEN experts. With just a few clicks, you can investigate where a gene is up- or down-regulated in human disease studies, and which statistical comparisons found differential regulation for *ACE2*. Land Explorer for IPA enables you to drill down on the studies that are of most interest, with filters for metadata. For example, if you are studying *ACE2* you may be interested only in projects concerning infectious diseases (Figure 5A). Because the QIAGEN curation teams carefully curate samples using a controlled vocabulary for each project, the metadata filters in Land Explorer for IPA let you quickly narrow down your focus to only the collection of studies that are of most interest. Finally, you can look for co-expression or co-regulation of genes like *ACE2* with other genes like *TMPRSS2*—one of the essential host factors for SARS-CoV-2 pathogenicity (Figure 5B).

Figure 5. (A) Comparison of *ACE2* gene expression by disease state. Statistical comparisons for differential expression of *ACE2* in infectious diseases (viral) using a bubble plot. X-axis represents differential gene expression of *ACE2* by Log2 fold change. Y-axis represents individual disease states. Bubble size represents the significance (p-value) from the differential expression (DESeq2 for RNA-seq and General Linear Model for microarray) of *ACE2*. Larger bubbles represent greater significance. Selecting dot(s) in this view will populate a Details table at the bottom of the view showing some key metadata from the comparison(s). **(B) Co-expression of *ACE2* and *TMPRSS2* in human tissues.** The multi-gene correlation view shows a scatter view comparing the RNA-seq data for *ACE2* and *TMPRSS2*. This scatter plot shows that *ACE2* and *TMPRSS2* are highly co-expressed in lung and brain samples. Values plotted are Log2(FPKM + 0.1).

Reveal the secrets of thousands of 'omics datasets to gain deeper biological insights

Land Explorer for QIAGEN IPA is a powerful resource for life science researchers who want to gain deeper insights into their biological data by quickly and easily leveraging the vast number of available 'omics datasets. It serves as a central location to visualize and compare these data in many different ways with just a few clicks. With a continually growing catalog of over 500,000 curated 'omics datasets, each which are carefully re-processed, you can quickly find answers to your most pressing research questions. Explore valuable 'omics datasets in a simple and interactive way in order to easily understand gene expression data across tissues, discover how genes are regulated in diseases and in response to treatments or mutations, and produce publication-quality visuals to communicate your findings.

Dr. Darbro reiterates, "I think the number one thing that stands out in QIAGEN IPA versus any other solution is that it's intuitive and easy to use. Number two is the sheer comprehensiveness. I use IPA quite a bit because there's nothing that has more information. There's nothing out there that has this kind of knowledge base behind it. The other thing that IPA has is other tools. It's great to have all these analysis tools just in one place. I don't think there's been a single project that I've worked on that we haven't found a use for some of the tools from IPA like Land Explorer." Dr. Darbro reflects, "I first became a QIAGEN IPA customer at least 10 years ago, and every single time we must make a decision whether to renew [the license], we have."

Understand complex 'omics data. Start your free trial of Land Explorer for QIAGEN IPA today.
Visit **digitalinsights.qiagen.com/IPA**

QIAGEN Ingenuity Pathway Analysis (QIAGEN IPA) is intended for molecular biology applications. This product is not intended for the diagnosis, prevention or treatment of a disease.

For up-to-date licensing information and product-specific disclaimers, see the respective Ingenuity product site. Further information can be requested from ts-bioinformatics.qiagen.com or by contacting your local account manager.

Ordering **www.qiagen.com/shop** | Technical Support **digitalinsights.qiagen.com/support** | Website **digitalinsights.qiagen.com**