



iRepertoire, Inc.

RepSeq & RepSeq+ Data Management Policies

For research use only; not to be used for clinical diagnostics.
Copyright © 2022 iRepertoire, Inc. All rights reserved.

V20230308



iRepertoire, Inc.

The iR logo is a registered trademark of iRepertoire, Inc. iRepertoire™, iRweb™ and the iPair Analyzer™ are trademarks of iRepertoire, Inc. Illumina®, iSeq®, MiSeq®, MiniSeq, HiSeq®, NextSeq®, and NovaSeq® are registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owner(s).

iRepertoire, Inc. does not assume any liability, whether direct or indirect, arising out of the application or use of any products, component parts, or software described herein or from any information contained in this manual. Furthermore, sale of iRepertoire, Inc. products does not constitute a license to any patent, trademark, copyright, or common-law rights of iRepertoire or the similar rights of others. iRepertoire, Inc. reserves the right to make any changes to any processes, products, or parts thereof, described herein without notice. While every effort has been made to make this manual as complete and accurate as possible as of the publication date, iRepertoire assumes no responsibility that the goods described herein will be fit for any particular purpose for which you may be buying these goods.



iRepertoire, Inc.

I. Overview

iRepertoire, Inc. (herein “iRepertoire”) provides products and services for sequencing the immune repertoire and single cell VDJ pairing information. In addition to these products and services, iRepertoire provides its Clients (herein “clients” or “user”) data output (herein “Output”) through its web-based analysis platform, iRweb™ or the iPair Analyzer™ (herein “Platforms”)

II. Research Use Only Data

Data generated from iRepertoire’s products and services are for Research Use Only and not to be used for clinical diagnostics unless stated otherwise.

III. Standard Data Processing Versus GDPR Data Processing

iRepertoire has enlisted two tiers of data control: standard data processing and EU General Data Protection Regulation (GDPR 2016/679). Standard data processing includes the processing of raw Illumina FASTQ files on iRepertoire’s data servers and further display on iRepertoire’s proprietary immune repertoire viewing tool, iRweb. In the European Union, certain data may be subject to the provisions of GDPR. A GDPR agreement must be in place for data to be processed under GDPR standards, and it is the client’s responsibility to ensure a Data Processing Agreement (DPA) is in place with iRepertoire prior to initiating services or data transfer so that iRepertoire is aware of the additional data handling responsibility. GDPR data is not available for viewing through iRweb since iRepertoire’s policies prohibit deletion of data from iRepertoire’s database as such action could lead to adverse events for other clients. For GDPR restricted data, raw and processed data files can be downloaded through iRepertoire’s secure file transfer protocol (SFTP). GDPR data is routinely deleted from iRepertoire’s servers, and it is the client’s responsibility to ensure that data is downloaded promptly. Once data is deleted from iRepertoire’s servers, we will no longer have access to the data files. Due to the level of security required to handle GDPR data, additional data handling fees apply as detailed in this document.

IV. Proprietary Information and Confidentiality

iRepertoire provides products and services for generating immune repertoire sequencing data. iRepertoire does not claim any reach-through rights to sequencing data produced with our technology. Project-specific intellectual property, as it pertains to sequencing data, may be governed by separate master services, material transfer, and/or non-disclosure agreements. In these cases, the terms of those agreements supersede this document. All proprietary information (including, without limitation, technologies, methodologies, primer sequences, and protocols) used to produce immune repertoire data remains the sole property of iRepertoire.



V. Decision-Making and Assessment of Analyzed Data

Any and all decisions or assessments made in relation to patterns or metrics present on iRweb are the sole responsibility of the user. iRepertoire assumes no legal responsibility for decisions based upon or assessments made with Output from Platforms. iRepertoire may provide insight and advice based on publicly-accessible literature and available experimental data upon customer request. These requests may be subject to the bioinformatic support fees detailed in Section VIII and summarized in **Table 1**.

VI. Complimentary Data Mapping and Repertoire Visualization via iRweb and the iPair Analyzer

iRepertoire provides basic data mapping analysis complimentary with all services provided. Bioinformatic analyses for data submitted following the use of any of iRepertoire's products is subject to the fees listed in **Table 2**. Bioinformatic analyses are output to a web-based software, iRweb. iRweb houses analyzed immune repertoire sequencing data that is secured by individual access credentials. The Output within iRweb is accessible only to authorized users for their specific analysis. iRepertoire maintains access to all accounts for troubleshooting purposes and customer support only. iRepertoire supports only those features provided in the iRweb platform (listed below). Additional information can be found in iRepertoire's **Data Analysis Guide**.

- Amino acid lists for unique CDR3s (uCDR3s)
- Diversity measurements, including Shannon Entropy, Diversity Index, and D50
- Unique CDR3 sharing across samples

Data for single-cell projects (iPair/ iPair+) are provided through email, Dropbox®, or a pre-determined location based upon User preference. This data can be viewed using a local application provided by iRepertoire called the iPair Analyzer™. Specific details regarding data output for single cell data are discussed in the **iPair Analyzer User Guide**.



VII. Access to Analyzed Data

1. Data Supplied Externally

- 1.1. iRepertoire waives liability and responsibility for the long-term or indefinite storage of data produced outside of its facilities with our primer systems. It is the responsibility of the customer to maintain storage, appropriate backup, and access to all data generated in laboratories and on sequencers outside of iRepertoire's facilities. This data may be shared with iRepertoire for the purpose of analyzing data, but iRepertoire has no obligation to store raw sequencing data files for the purpose of a secondary data backup site.
- 1.2. Intermediate processed data files such as demultiplexed files, stitched reads, and various stages of the analytic pipeline, are routinely deleted from the server; however, data stored on iRweb remains intact indefinitely and is available 24 hours per day 7 days per week (except during routine maintenance). Intermediate files can be shared with customers free-of-charge in the 2 month period following notification of the release of analyzed data. Access to this data must be requested in writing through email or on the Data or Service Submission Form(s). After the initial 2 month period, bioinformatics fees for data handling and management and a flat \$300 re-processing fee per flow cell or lane of data will be levied in order to access intermediate data after the 2 month's notification time frame.

2. Data Generated Through Service Projects

- 2.1. In the event that User libraries are placed in communal pools (e.g., in sequencing runs with other customer libraries or with internal iRepertoire R&D libraries), only demultiplexed data can be provided to customers. This data is sample-specific and extracted via either our proprietary molecular barcoding system or the Illumina® dual-index system, depending upon products utilized in service. This system provides confidentiality and maintains access to data only by authorized parties.
- 2.2. In the event that a customer purchases all sequencing space within either a flow cell or lane (as per Illumina® product definitions), raw FASTQ files output by the sequencer can be shared directly with the customer. Sharing can be performed through multiple channels (e.g., Basespace®, FTP/SFTP, Dropbox®, etc.), but methods must be discussed prior to analysis.
- 2.3. All MiSeq® runs produced as a part of service are automatically stored within iRepertoire's own Basespace accounts and are therefore subject to all terms and conditions that apply. If the terms and conditions provided for Illumina's Basespace accounts are not agreeable, iRepertoire should be notified prior to sequencing so that the sequencing instruments can be disconnected from Basespace and stored locally. The latter operation can only be performed when sequencing flow cells are purchased outright. iRepertoire does not utilize Illumina Basespace for NextSeq and NovaSeq platforms. All three platforms are processed on internal servers after sequencing and processed data is released through iRepertoire's iRweb (unless data is GDPR-restricted).
- 2.4. ***For GDPR restricted data, sequencers are disconnected from the internet, and data is transferred locally within the network. GDPR data cannot be provided through Basespace or iRweb.***



2. Data Generated Through Service Projects (Continued)

- 2.5. All HiSeq®, NextSeq®, or NovaSeq® run data or any GDPR-restricted data are shared through either iRepertoire's SFTP server or hard drive shipment (not Basespace). Hard drive shipments will incur fees including purchase of the hardware and shipping and handling fees, which are both dependent upon project size. These fees will be established through official quote and purchase order (PO) prior to shipment.
- 2.6. Raw (FASTQ files directly from sequencer) or demultiplexed sequencing data are held in temporary local storage for at least three months. Non-GDPR raw data files are placed in long-term storage both locally and through AWS® cloud storage services.
- 2.7. Intermediate processed data files such as demultiplexed files, stitched reads, and various stages of the analytic pipeline, are routinely deleted from the server; however, data stored on iRweb remains intact indefinitely. Intermediate files can be shared with customers free-of-charge in the 2 month period following notification of the release of analyzed data. Access to this data must be requested in writing through email or on the Data or Service Submission Form(s). After the initial 2 month period, bioinformatics fees for data handling and management and a flat \$300 re-processing fee per flow cell or lane of data will be levied in order to access intermediate data after the 2 month's notification time frame.
- 2.8. For standard data, iRepertoire will store copies of data produced from service projects indefinitely unless removal of such data is specifically requested. Requests to remove raw data from backups will be assessed on a case-by-case basis and will be subject to data handling fees at \$200 per hour. For libraries pooled with other projects (without the use of Illumina® dual indices), removal of raw data files is not permissible. For GDPR data promulgated by a separate DPA, raw and processed data is deleted on a quarterly basis after notification by iRepertoire.
- 2.9. ***For GDPR data, it is the customer responsibility to ensure all raw files have been downloaded and appropriately stored prior to deletion and that the data fidelity is maintained post-download. Once GDPR data is erased, iRepertoire will no longer have access to any of the project related data files.***



VIII. Bioinformatic Support

1. Initial teleconference-based consultations regarding study design (as it relates to read depth and sample coverage), sample requirements, and the initial release of analyzed data are offered as part of our standard service. Bioinformatic consultations are available via teleconference if requests for analyses exceed what is currently offered through the basic data analysis package. Bioinformatic fees are assessed at \$200 per hour.
2. Included with product and service projects is an optional 1 hour walk-through of the features of iRweb, via teleconference, which is offered in addition to the **Data Analysis Guide**. One session is included per Principal Investigator’s laboratory per year. Additional training is subject to hourly bioinformatic fees.
3. In addition to all of the tools described in **section VI**, iRepertoire provides a ‘Raw Data’ download in iRweb in which the majority of the underlying mapped data used to generate the graphics in iRweb are available in .csv formats.
4. All graphical portions of the iRweb analyses are accessible to the customer and can be downloaded individually from sample to sample. iRepertoire does not provide mass download of graphics for projects in excess of 10 libraries.
5. High-resolution versions of Tree Maps, graphical representations of CDR3 diversity, can be provided at no cost within 2 months of the notification for release of data. After this 2 month period, Tree Maps can be provided at a cost of \$15 per high resolution Tree Map provided.

Table 1. Bioinformatics and Data Management Fee Table

GDPR Data Handling	\$600 per flow cell
Raw sequencing data retrieved from long-term (6+ month) storage	\$180 per project or flow cell \$500 for NovaSeq projects
Re-processing fee*	\$300 per flow cell (under 40 GB)*
Bioinformatic fees	\$200 per hour
HD Tree Maps	\$15 per image provided

iRepertoire reserves the right to adjust fees at any time for any reason after evaluation of request. These adjustments will be first discussed with User.

**Re-processing fees for data over 40 GB will be provided by quotation after assessment by the iRepertoire computational and bioinformatics team.*



IX. Bioinformatic Analysis of Submitted Data

1. Our bioinformatic pipelines for RepSeq and RepSeq+ are built around the industry standard FASTQ file format. **The preferred data submission file format is FASTQ.gz. The only two file formats we can accept are FASTQ and BCL* files. BCL files are not accepted for NovaSeq data.**
2. The submission of BCL files requires that the entire Run Folder* is shared with us for processing, and the conversion of these files to a FASTQ format **will** incur data management fees as detailed below. *We do not accept BAM, SAM, FASTA, BED, VCF, or other file formats.*
3. Deep sequencing of libraries can result in large FASTQ files per library. These special cases of library sequencing can lead to issues with data transfer and processing. For deeply sequenced libraries (ie, one dual index pair >10GB), additional data overage fees apply and will be provided by quotation after assessment by the iRepertoire computational and bioinformatics team prior to the start of analysis.

Table 2. Bioinformatic Analysis Fees	Price per Gigabyte	Notes
File Manipulations*	\$15	Includes but not limited to: bcl2FASTQ and bcl2FASTQ2 conversion per flow cell or lane; re-compression or concatenation of files; the processing of ≥ 10 demultiplexed libraries.
	Price per Gigabyte Per FASTQ	
RepSeq Data Analysis	Complimentary	First 40 GB of data are processed at no charge (per flow cell)
RepSeq+ Data Analysis	Complimentary	First 100 GB of data are processed at no charge (per flow cell)
Data Overages	\$15	Over >40GB for RepSeq and over 100 GB for RepSeq+
Priority Analysis Fees	\$50 per Gigabyte	

*BCL folder

1. The BCL folder is the Illumina output run directory containing the BCL files from a sequencing run. All files and folders within the BCL folder must be untouched when sent to iRepertoire for analysis.
2. When sending the BCL folder to iRepertoire for analysis, the BCL folder may be given to iRepertoire through either the SFTP server or Google Drive. If placed on Google Drive, the BCL folder must be compressed in either ZIP format (.zip) or as a GZIP TAR file (.tar.gz). Failure to provide BCL folders in the correct format will result in analysis delays.
3. The BCL folder will be converted to FASTQ files for analysis.
4. iRepertoire reserves the right to define “File Manipulations” as any task beyond the analysis of standard FASTQ-formatted file submissions.

FASTQ file

iRepertoire’s analysis pipelines are based around the FASTQ format. FASTQ files sent to iRepertoire are expected to be compressed in GZIP format (.FASTQ.gz).



IX. Standard Turnaround Times For Standard Analysis And Priority Analysis

1. We allow for up to three weeks for the analysis and release of all data analyzed by our pipeline. This is to account for the traffic of analyzing smaller iSeq 100 data sets (≤ 1 GB in size) to much larger data sets like those produced by the NovaSeq and NextSeq systems (which can exceed 200 GB in total size). Data analysis is currently handled in a “first come, first serve” basis, with the exception of service projects that have paid a Rush fee. The Rush fee for service projects places the customer’s data analysis at the head of the current queue.
2. Priority Analysis fees are assessed at \$50 per Gigabyte per flow cell or lane of data. These fees are applicable to customers that have performed sequencing in their own labs and would like be placed in the front of the analysis queue. This fee is similar to the Rush fee applied to service projects and does not guarantee a turn around time, but does ensure the data is analyzed next in the queue. *For example, for priority analysis of a MiSeq run with 6 GigaBytes of data total (3GB R1 and 3GB R2 or 6 GB SER), the total priority analysis fee would be assessed as \$50 x 6 or \$300 for the flow cell priority analysis.*

In order to be able to provide Priority Analysis, there are very specific rules and procedures that *must* be adhered to:

- 2.1. Customers **must** send a notification email to DataSupport@iRepertoire.com upon the completion of the upload of their data to our SFTP server or upon the sharing of a link to download the files. Unless a notification email is sent, the Priority Analysis service **will not** be applied to the submission.
- 2.2. If demultiplexed or stitched files are requested for the Priority submission(s), the notification email sent to DataSupport@iRepertoire.com **must** include this request in order to be honored.
- 2.3. The Priority Analysis fee **does not** guarantee a given turnaround time.
- 2.4. Notification of the Priority Analysis **must occur via email notification** and **cannot** be included in the comments section of Data Submission Forms.