General Terms and Conditions

Validity of the offer

This offer is valid from its date of issue until the 60th calendar day thereafter. Upon expiry of the aforementioned period, unless IGA Technology Services srl receives formal acceptance from the Customer, the Offer will be considered to have expired.

Orders

The Customer must send the order to IGA Technology Services srl before or, at the latest, at the same time as the samples, by email (to the address orders@igatechnology.com or fax to No. 0432 603887)

Samples

Samples should be sent together with the compiled Samples Spreadsheet providing the quote reference. Include the printout with the shipped parcel and send a copy via e-mail. To be able to properly track and safeguard your samples please send us the Tracking Number via e-mail when available

The samples must meet the quality criteria required in the instructions attached to the Offer. The Customer will be alerted to any non-compliance and must decide whether to carry out the service anyway or to replace the non-compliant samples. In the event the Customer decides to proceed with the service anyway, the total or partial failure of the experiments will be entirely the responsibility of the Customer and will therefore be invoiced at the full price.

In the event the supplied sample is only partially used its restitution can be requested by and no later than 30 days from delivery of the results.

Communications

All requests for information about the service must only be sent via email (orders@igatechnology.com) and will be dealt with within 2 working days from receipt. In the event problems arise during the processing, the Customer will be contacted directly to agree on possible solutions. Otherwise, the Customer will only be contacted at the end of the service.

Invoicing

For orders of net amounts less than 10,000 EUR, invoicing will take place at the end of the service. For orders of net amounts greater than 10,000 EUR, invoicing will take place in two stages:

- 50% on receipt of the samples;

- 50% at the end of the service.

Should the service not be successful due to technical problems, it will be repeated without any additional charges to the Customer. Should the service not be successful on two consecutive occasions, 50% of the amount of the order will be invoiced in any case to partially cover the costs sustained.

Delivery times

Our standard data delivery time is between 6 and 10 weeks, based on the type of service required, the need of specific data analysis and the optimization of the use of the Illumina platform, as described below. In any case, we will guarantee delivery of the data within maximum 16 weeks from the date of receipt of the samples.

For special projects in terms of sample number (e.g. thousands of samples for Target Genotyping by Sequencing or ddRAD genotyping) the time for data delivery will be agreed with the client on a project basis.

Optimization of the use of the Illumina platform

A flowcell consists of more than one lane (2 or 8, depending on run type) and each lane can accommodate one or more samples depending on the level of multiplexing/reads number chosen by the Customer. Should the service require only the partial use of a flowcell, the samples will be processed when the remaining capacity of the flowcell is taken up by the samples of other Customers. For this reason, the delivery times given in the Offer could undergo modifications (maximum 16 weeks).

In any case, IGA Technology Services srl will not be responsible for delayed deliveries or services cancelled due to *force majeure* or, in any event, factors beyond its control, such as technical problems with the instrumentation, delays in the delivery of the reagents by the suppliers or samples provided in formats or quality that do not meet our application-specific requirements.

Method of delivery and data storage

At the end of the service, the data will be made available to the Customer on a server of IGA Technology Services srl. The Customer will receive a username and personal password to access the data, which must be downloaded from the server within 15 days from receipt of the notification of availability.

The Customer may request the delivery of the data saved onto an external hard-disk, sustaining the associated additional costs.

The data will be retained on a server of IGA Technology Services srl for **3 months, starting from the date of delivery**. After this period, data will be deleted without notification. If extended persistence is required, it must be notified beforehand (within 2 months from data

delivery). Extra charge will be applied for extended storage. Upon expiry of the 3 months, the data will be completely deleted. Therefore, once this period has passed, any requests to upload the data on the download service server cannot be met.

Data download and integrity check

After notification, data will be available for download for a period of two weeks. After this period of time, in case data is no longer available in the download page, a request must be issued to reload the data on the export server. After download, customer is responsible of checking data-integrity by the provided MD5 files. If integrity check fails, please retry the download. If data integrity issues persist, please communicate promptly to IGATech, before the 3 months' storage period terminates. To perform data integrity-check on a Windows system, download the software MD5Checker (http://getmd5checker.com/), once installed, start the software and load (Edit → Add/Open) the provided MD5 file, the integrity checking will start automatically. Please ensure that the MD5 file resides at the same folder level as provided in the download page.

Confidentiality

All customer information is held strict confidence. All materials and information sent to us and the data produced by us for the order are the proprietary of the customer and will be returned to the customer or discarded in confidential manner. We archive customer materials only when instructed to.

Standard Bioinformatics Services and Outputs

All the standard analyses include the delivery of raw data or trimmed data (quality filtered) - please communicate in advance which type of data is desired. When not specified raw data will be delivered. Raw data sequences are processed by our internal pipeline to have already any adapter read-through masked by N characters. Quality control reports will be provided for each reads set.

Common file formats references

FASTQ: https://en.wikipedia.org/wiki/FASTQ_format

SAM/BAM: https://samtools.github.io/hts-specs/SAMv1.pdf

GTF/GFF: http://mblab.wustl.edu/GTF22.html

VCF: http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-40

BED: https://genome.ucsc.edu/FAQ/FAQformat.html

BIOM: http://biom-format.org/

RNA-Seq (expression analysis)

- FASTQ.GZ files containing raw sequences.
- A file containing the estimated gene-level expression values with the FPKM read counts and normalized read counts (gene_abundance.tab).
- A series of metrics files describing overall quality metrics computed from the BAM file.
- A file in BedGraph format allowing for data visualization in a Genome Browser such as UCSC
- A PDF file describing the analysis flow (Customer_mRNASeq-analysis_report.pdf).
- Differential expression analysis output (only when required), detailed description of each file is in "Customer_mRNASeq-analysis_report"; values refer to post-normalization/scaling procedures:
    - 100-most-expressed-genes_clust.pdf is a graphical representation of the clustering of 100 most expressed genes overall
    - DE_gene_list is the list of significantly differentially expressed genes
    - whole_dataset.variance_stab_transf.tsv is a table containing all data after variance stabilization transform (VST) as described in DESeq2
    - DE_genes.variance_stab_transf.tsv same as above but limited to genes found to be differentially expressed in any of the tests
    - dipsersion_fit.pdf depicts dispersion of the data over the mean of the expression and the curve that has been fitted to perform GLM testing
    - [test_control]*_res_ordered.csv is a comma-separated table with all the statistical testing for a given pairwise test
    - [test_control]*_res_ordered. alpha0.01_DE.tsv same as above but limited to genes with an adjusted p-value less than 0.01
    - [test_control]MAplot.pdf MA-Plot for a given test
    - VST_PCA.pdf A PCA analysis of all samples/groups with VST-normalized data
    - VS T_samples-distances.pdf clustering of samples using VST-normalized data.
    - Files for hierarchical and k-means cluster analysis visualization (can be visualized with JavaTreeView,).

IMPORTANT: For RNA-seq analyses, we strongly advocate the presence of a minimum of three biological replicates for each test/control. With no replicates, false negatives in the detection of differentially expressed genes are common. The standard analysis is performed using best practices using STAR aligner, StringTie and DESeq2 R package for a standard "test-vs-control" significance test. Models that are more complex (with interaction of several factors or uncontrolled "batch effect") can be supported as well: please inquire. We also encourage our customers to perform the most of desired experiments (library preparation at least) at once, reducing costs and eliminating batch effects which sometime can mask truly differentially expressed genes by an increased within-replicates variance.

RNA-Seq (de novo assembly)

- A FASTA file with the entire set of reconstructed transcript fragments
- A report file describing summary metrics of the assembly
- *[when requested]* A table of functional annotation with the assigned Geno Ontology terms for each sequence

DNA-Seq de novo assembly

- A FASTA file with the entire set of reconstructed contigs
- A FASTA file with the entire set of reconstructed scaffolds
- A report file describing summary metrics of the assembly
- [*when requested*] A GFF file with predicted gene models
- [*when requested*] A table of functional annotation with the assigned Geno Ontology terms for each gene model (via InterproScan5)

ddRAD genotyping

- A catalog file reporting all reference loci (with the genomic coordinates if reference-based analysis) with a representative consensus sequence. Positions (if alignment-based) always refer to the left-most coordinate of the RAD locus as represented.
- A table reporting all SNP sites found with respect to catalog loci and the position within such.
- A table reporting all possible haplotypes found with respect to catalog loci.
- Raw de-multiplexed and uniformed data (i.e. 1 or 2 fastq file(s) per sample, with internal barcode removed, all uniformed to the same length)
- [*for natural population*] A VCF file with the population-wise single-SNP genotype calls
- [*for natural population*] A VCF file with the population-wise haplotype calls
- [*for natural population*] Data converted in the *structure* format
- [*for natural population*] Data converted in the *genepop* format
- [*for natural population*] Data converted in the *plink ped/map* format
- [*for natural population*] A table reporting all raw (UNFILTERED) haplotypes
- [*for bi-parental crosses*] A segregation matrix in the *joinmap/onemap* format (if reference-based analysis, markers IDs will be converted to genomic positions)

Target-enrichment for genotyping (hybridization-capture/single-primer enrichment)

- FASTQ.GZ files containing raw sequences.
- A BAM file containing high quality reads aligned to the reference sequence.
- Quality control of reads
- Enrichment report (amount of reads mapped on target regions)
- A VCF file reporting multi-sample SNP/DIP calling
- [*when requested*] Functional annotation of variant effect (Syn/non-Syn/nonsense) and position with respect to the closest gene.

Metagenomics for 16S/18S/ITS

- A FASTA file with all the filtered sequences used during the analyses (i.e. after pair-overlap, quality trimming and chimera removal); the header of each sequence will contain sample name.
- A FASTA file with a single sequence chosen to be representative of an OTU (either being referred to a database or constructed by clustering)
- A map of OTUs with all the clustered sequences.
- OTU tables at each taxonomical level (phylum, class, order, family, genus, and species) with relative abundance for each sample in the cohort. OTU tables are provided in both tab-separated files and BIOM files.
- A bar chart plotting accompanied by colored tables with relative taxonomical abundance of each samples in html format.
- Rarefaction analysis to measure alpha-diversity of each sample (observed OTUs, Chaos1, Shannon index, Simpson index), with plots (can be interactively colored with additional metadata provided by customer before ahead)
- PCoA analyses to measure beta-diversity via Bray-Curtis dissimilarity

IMPORTANT: to maximize analyses information, please provide any possible categorization of samples. This will allow navigating the data with more insights and providing useful statistical testing.

Metagenomics for custom amplicons (pair-overlap)

- A FASTA file with all the filtered sequences used during the analyses (i.e. after pair-overlap, quality trimming and chimera removal); sequences are “de-replicated”, meaning that a single sequence is chose as representative for a set of identical reads; the header of such sequences will contain sample name and the number of identical reads present in the raw data.
- A FASTA file with a single sequence chosen to be representative of an OTU
- OTU table (TSV file) with relative abundance for each sample in the cohort.
- Rarefaction analysis to measure alpha-diversity of each sample (observed OTUs, Chaos1, Shannon index, Simpson index), with plots (can be interactively colored with additional metadata provided by customer before ahead)
- PCoA analyses to measure beta-diversity via Bray-Curtis dissimilarity
- A Blast-based (lowest common ancestor) taxonomy assignment table to each OTU.

Metagenomics WGS (basic Kraken search)

- A table summarizing relative abundance of each taxonomy level assigned to reads.
- An interactive, hierarchic, visualization of abundance across samples made with a *krona* pie (https://github.com/marbl/Krona/wiki).

Whole genome / Exome sequencing (variant calling and annotation)

- FASTQ.GZ files containing raw sequences.
- A BAM file containing high quality reads aligned to the reference sequence.
- A PDF file describing quality metrics computed from the BAM file.
- An XLSX file reporting all variants (SNVs and DIPs) with amino acid change annotation, functional annotation (for Whole Genome Sequencing) and other information that are helpful when prioritizing variants.
- A VCF file containing all variants (SNVs and DIPs), with a filter flag for each variant.
- A PDF file describing the analysis flow.

smallRNA (human only)

- FASTQ.GZ files containing raw sequences
- A PDF file reporting the histogram of size distribution and the classification of small RNA sequences
- CSV files (miRNA.csv and piRNA.csv) containing counts of tags mapping against all miRNAs available in mirBASE and piRNAs available in piRNABase. Results are provided as raw counts.
- Differential expression outputs (only when required):
    - A tabular file containing list of miRNAs with tag-counts per Sample1 and Sample2 (including each replicate) and fold changes, p- and q-values.
    - Differential expression MA-plot, which represents the log2 fold change (M) versus the mean (A) of normalized count between the two samples.

CHIP-Seq

- FASTQ.GZ files containing raw sequences
- BAM file containing reads aligned to the reference sequence.
- peaks.xls: a tabular file containing list of peaks and their length, summit location and height, as well as their fold change, p- and q-value.
- narrowpeak.bed: List of peak locations, q-values, fold change, p-value, q-value (again) and summit position relative to peak start, in narrow peak format (BED6+4).
- summits.bed: List of peak summits and q-values in BED format.
- model.pdf: If the peak model building is successful, a plot of the model is generated. The shape of the modeled peaks allows to assess the quality of the model.
- Peak annotation in “peaks.pdf” file displaying ChIP peak distribution across functionally important genomic regions, average enrichment signal within/near genes and the peaks' intensity distributions across chromosomes. Available only for somegenomes (ce6 for worm, dm3 for fly, mm9 and mm10 for mouse and hg19 for human).
- “Table.stat” on read distribution across different species in order to test for contaminates.

- Table with read distribution over different categories (mitochondrial, ribosomal, chloroplast, etc.) in “specific_categories.stat”.
- A PNG plot (“fingerprint.png”) reporting cumulative coverage on ordered windows. along the genome. This plot is diagnostic of coverage uniformity: it is expected to be close to a diagonal in control samples and with a rapid slope in IP-samples (steepness depends on target type and enrichment efficiency).

Bioinformatics Support Services

We have divided our portfolio of bioinformatics services into four different levels of support: from the most standard established workflows to the more customized data manipulation and interactive customer-analyst follow-up toward data refinement. The document also describes the standard result bundles ("BRONZE") for some of our most consolidated pipelines. We also provide conditions of data storage, download and checking. Please read this document carefully.

- **BRONZE Level ("Standard output")**

  Customer will receive analytical results in standard formats as described for each standard application in the paragraph below (e.g. BAM, FASTA, FASTQ, VCF, etc.) or as defined in a specific quote. These results are obtained from standard validated pipelines of IGATech and do not account any further manipulation at this level. IGATech Bronze includes the computation time for one single run of the analysis without extra support, but for data download and understanding of result files.

- **SILVER Level ("Data conversion - custom graphics ")**

  Customer will obtain transformed data (conversion of formats) or visualization products not regularly produced by our internal pipelines or by software already validated in our production pipelines. Custom conversion scripts will be developed *ad-hoc* to obtain data transformation/visualization. IGATech will maintain rights on the property and use of such software for other purposes (software products are not part of deliverables, unless otherwise agreed). Terms and costs of these services will be evaluated on a case-by-case basis prior to the order. Costs will depend on the hands-on time required by the bioinformatics staff. IGATech Silver does not include development of software for data analysis, computation or statistical inference.

- **GOLD Level ("Custom Analysis Support A")**[1]

  Cutomer will receive customized support on i) experimental design, ii) development of dedicated software for data analysis or statistics, iii) testing of new software/approaches not already accounted in our standard services, iv) custom data integration from different sources. IGATech will maintain rights on the property and use of such software for other purposes (software products are not part of deliverables, unless otherwise agreed). The main objectives, terms and costs of the study must be agreed before analysis kick-off. The completion of the analyses must be feasible in a time span of **2 months** after data production or data upload (in the case of data from third parties). The active hand-on time of bioinformatics staff will not exceed **60h**.. Analyses can be repeated when it is demonstrated that previous ones have generated incorrect results or partial results for reasons due to technical issues, incomplete data sets or wrong parameters; this does not include cases when incomplete results or due to bad starting material or wrong experimental design. Repeats cannot exceed the maximum hands-on time. Time to set up extra analyses, phone calls and meetings are counted as hands-on time, while CPU time is not counted. This level includes up to

one extra cycle of intensive analyses[2] requested by the customer (unlike aforementioned repeats), regardless of the reason of the request (e.g. to change parameters, reference database or software for publication purpose).

- **PLATINUM Level ("Custom Analysis Support B")[1]**

  Customer will receive customized support either in i) experimental design, ii) development of dedicated software for data analysis or statistics, iii) testing of new software/approaches not already accounted in our standard services, iv) custom data integration from different sources. IGATech will maintain rights on the property and use of such software for other purposes (software products are not part of deliverables, unless otherwise agreed). The main objectives, terms and costs of the study must be agreed before analysis kick-off. The completion of the analyses must be feasible in a time span of **6 months** after data production or data upload (in the case of data from third parties). The active hand-on time of bioinformatics staff will not exceed **120h**. As a major difference from the GOLD package, during this period, Customer will have the opportunity to interact with the analysts to have feedback on the proceeding of experiments and drive downstream analyses refinement. This will be carried based on step-by-step inspection of preliminary data and analysis results during the first 4 months of the reserved time; the last two months are reserved to IGATech to complete development and analyses as agreed in advance. Analyses can be repeated when it is demonstrated that previous ones have generated incorrect results or partial results for reasons due to technical issues, incomplete data sets or wrong parameters; this does not include cases when incomplete results or due to bad starting material or wrong experimental design. Repeats cannot exceed the maximum hands-on time. Time to set up extra analyses, phone calls and meetings are counted as hands-on time, while CPU time is not counted. This level includes up to two extra cycles of intensive analyses[2] requested by the customer (unlike aforementioned repeats), regardless of the reason of the request (e.g. to change parameters, reference database or software for publication purpose).

[1]On the basis of project complexity and customer requests a dedicated quote with presumptive man/hour will be issued. If circumstance will resolve all the analyses with a lower amount of time, the remainder hours will be accounted for further custom analyses or converted to standard bioinformatics services.

[2]Intensive analyses: computationally intensive tasks such as short reads alignments, SNP calling, *de novo* assembly, reads clustering or any task which can require more than 6 CPU/h.