UNIVERSITY OF LIVERPOOL | CENTRE FOR GENOMIC RESEARCH

DOWNLOAD PAGE FOR LIMS28866

Quick links

Data Download

- Trimmed data, suitable for most downstream analyses. See below for details of the trimming pipeline.
- Raw data. Statistics only. Please note that we do not routinely provide access to the untrimmed sequence files, since we recommend that the trimmed data are used for most downstream analyses.

To download multiple files we recommend using wget. This is likely to be already installed on Linux, and is available for Windows and Mac. To download all the trimmed data files use the command:

```
wget -r --cut-dirs=2 -np -nH -R "index.html*" https://cgr.liv.ac.uk/illum/LIMS28866_1d518edb8da77ccc/Trimmed/
```

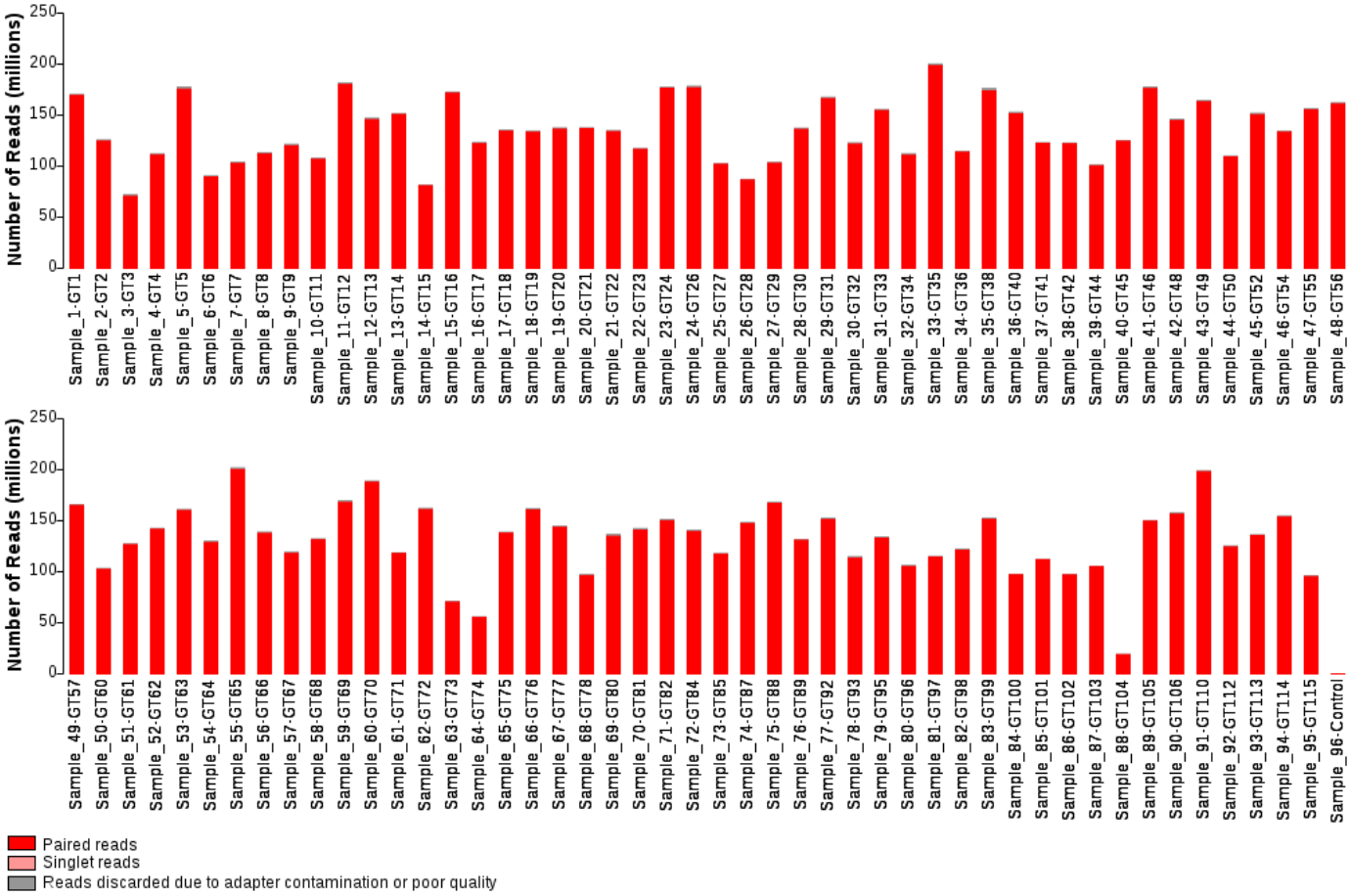
This will create a directory "Trimmed" containing all the .fastq.gz files. Be sure to copy the command exactly and include the trailing "/".

The sequence files are in .fastq.gz format. This is a version of the text FASTQ format, which has been binary compressed using gzip to reduce the file sizes. Many analysis programs such as SPAdes and BWA can read .fastq.gz files directly, however for some software packages you may need to decompress the files using gunzip or 7-zip.

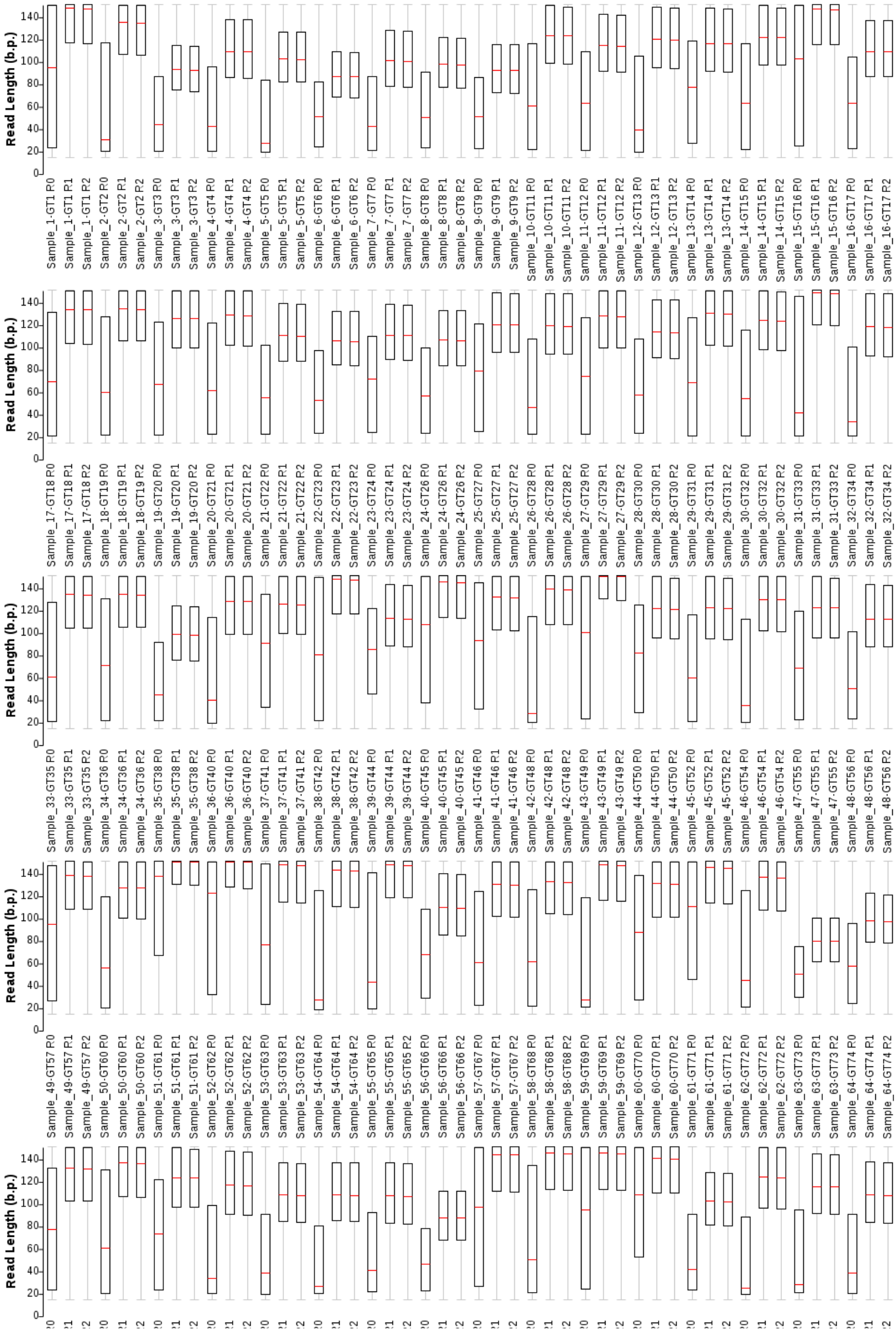
For paired-end sequence data, there are three sequence file types. The files labelled R1 and R2 contain the corresponding paired-end sequences. The singlet files contain sequences whose pair has been removed due to poor sequence quality or adapter contamination. If a sample has been sequenced several times, there will be several sets of sequence files in the sample directory. These will need to be concatenated before downstream analysis.

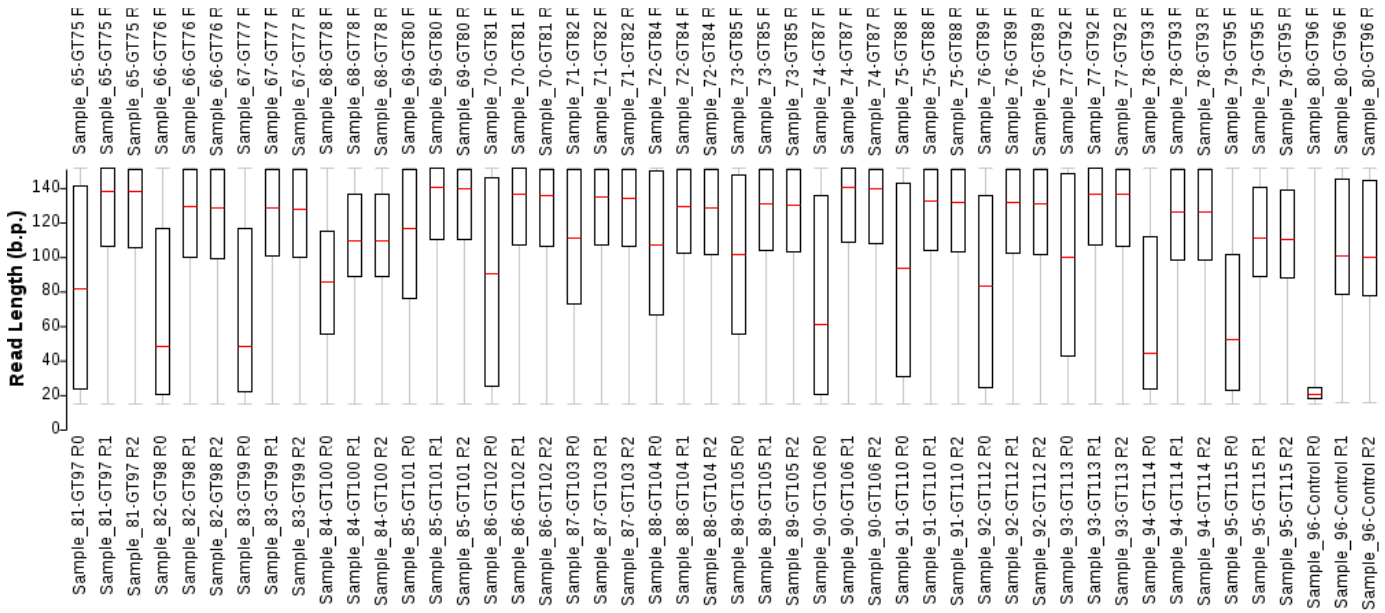
Summary Statistics

Diagram illustrating the total number of reads obtained for each sample.



Box plot showing the distribution of trimmed read lengths for the forward (R1), reverse (R2) and singlet (R0) reads. Note that it is common for a small number of reads to consist of mostly adapter-derived sequence, so it is expected that the distribution will show a long tail.


Read Length (b.p.)
0
20
40
60
80
100
120
140
Sample_1-GT1 R0
Sample_1-GT1 R1
Sample_1-GT1 R2
Sample_2-GT2 R0
Sample_2-GT2 R1
Sample_2-GT2 R2
Sample_3-GT3 R0
Sample_3-GT3 R1
Sample_3-GT3 R2
Sample_4-GT4 R0
Sample_4-GT4 R1
Sample_4-GT4 R2
Sample_5-GT5 R0
Sample_5-GT5 R1
Sample_5-GT5 R2
Sample_6-GT6 R0
Sample_6-GT6 R1
Sample_6-GT6 R2
Sample_7-GT7 R0
Sample_7-GT7 R1
Sample_7-GT7 R2
Sample_8-GT8 R0
Sample_8-GT8 R1
Sample_8-GT8 R2
Sample_9-GT9 R0
Sample_9-GT9 R1
Sample_9-GT9 R2
Sample_10-GT11 R0
Sample_10-GT11 R1
Sample_10-GT11 R2
Sample_11-GT12 R0
Sample_11-GT12 R1
Sample_11-GT12 R2
Sample_12-GT13 R0
Sample_12-GT13 R1
Sample_12-GT13 R2
Sample_13-GT14 R0
Sample_13-GT14 R1
Sample_13-GT14 R2
Sample_14-GT15 R0
Sample_14-GT15 R1
Sample_14-GT15 R2
Sample_15-GT16 R0
Sample_15-GT16 R1
Sample_15-GT16 R2
Sample_16-GT17 R0
Sample_16-GT17 R1
Sample_16-GT17 R2
Sample_17-GT18 R0
Sample_17-GT18 R1
Sample_17-GT18 R2
Sample_18-GT19 R0
Sample_18-GT19 R1
Sample_18-GT19 R2
Sample_19-GT20 R0
Sample_19-GT20 R1
Sample_19-GT20 R2
Sample_20-GT21 R0
Sample_20-GT21 R1
Sample_20-GT21 R2
Sample_21-GT22 R0
Sample_21-GT22 R1
Sample_21-GT22 R2
Sample_22-GT23 R0
Sample_22-GT23 R1
Sample_22-GT23 R2
Sample_23-GT24 R0
Sample_23-GT24 R1
Sample_23-GT24 R2
Sample_24-GT26 R0
Sample_24-GT26 R1
Sample_24-GT26 R2
Sample_25-GT27 R0
Sample_25-GT27 R1
Sample_25-GT27 R2
Sample_26-GT28 R0
Sample_26-GT28 R1
Sample_26-GT28 R2
Sample_27-GT29 R0
Sample_27-GT29 R1
Sample_27-GT29 R2
Sample_28-GT30 R0
Sample_28-GT30 R1
Sample_28-GT30 R2
Sample_29-GT31 R0
Sample_29-GT31 R1
Sample_29-GT31 R2
Sample_30-GT32 R0
Sample_30-GT32 R1
Sample_30-GT32 R2
Sample_31-GT33 R0
Sample_31-GT33 R1
Sample_31-GT33 R2
Sample_32-GT34 R0
Sample_32-GT34 R1
Sample_32-GT34 R2
Sample_33-GT35 R0
Sample_33-GT35 R1
Sample_33-GT35 R2
Sample_34-GT36 R0
Sample_34-GT36 R1
Sample_34-GT36 R2
Sample_35-GT38 R0
Sample_35-GT38 R1
Sample_35-GT38 R2
Sample_36-GT40 R0
Sample_36-GT40 R1
Sample_36-GT40 R2
Sample_37-GT41 R0
Sample_37-GT41 R1
Sample_37-GT41 R2
Sample_38-GT42 R0
Sample_38-GT42 R1
Sample_38-GT42 R2
Sample_39-GT44 R0
Sample_39-GT44 R1
Sample_39-GT44 R2
Sample_40-GT45 R0
Sample_40-GT45 R1
Sample_40-GT45 R2
Sample_41-GT46 R0
Sample_41-GT46 R1
Sample_41-GT46 R2
Sample_42-GT48 R0
Sample_42-GT48 R1
Sample_42-GT48 R2
Sample_43-GT49 R0
Sample_43-GT49 R1
Sample_43-GT49 R2
Sample_44-GT50 R0
Sample_44-GT50 R1
Sample_44-GT50 R2
Sample_45-GT52 R0
Sample_45-GT52 R1
Sample_45-GT52 R2
Sample_46-GT54 R0
Sample_46-GT54 R1
Sample_46-GT54 R2
Sample_47-GT55 R0
Sample_47-GT55 R1
Sample_47-GT55 R2
Sample_48-GT56 R0
Sample_48-GT56 R1
Sample_48-GT56 R2
Sample_49-GT57 R0
Sample_49-GT57 R1
Sample_49-GT57 R2
Sample_50-GT60 R0
Sample_50-GT60 R1
Sample_50-GT60 R2
Sample_51-GT61 R0
Sample_51-GT61 R1
Sample_51-GT61 R2
Sample_52-GT62 R0
Sample_52-GT62 R1
Sample_52-GT62 R2
Sample_53-GT63 R0
Sample_53-GT63 R1
Sample_53-GT63 R2
Sample_54-GT64 R0
Sample_54-GT64 R1
Sample_54-GT64 R2
Sample_55-GT65 R0
Sample_55-GT65 R1
Sample_55-GT65 R2
Sample_56-GT66 R0
Sample_56-GT66 R1
Sample_56-GT66 R2
Sample_57-GT67 R0
Sample_57-GT67 R1
Sample_57-GT67 R2
Sample_58-GT68 R0
Sample_58-GT68 R1
Sample_58-GT68 R2
Sample_59-GT69 R0
Sample_59-GT69 R1
Sample_59-GT69 R2
Sample_60-GT70 R0
Sample_60-GT70 R1
Sample_60-GT70 R2
Sample_61-GT71 R0
Sample_61-GT71 R1
Sample_61-GT71 R2
Sample_62-GT72 R0
Sample_62-GT72 R1
Sample_62-GT72 R2
Sample_63-GT73 R0
Sample_63-GT73 R1
Sample_63-GT73 R2
Sample_64-GT74 R0
Sample_64-GT74 R1
Sample_64-GT74 R2



Red line indicates median length

Box indicates interquartile range

Whiskers indicate minimum and maximum read lengths

Further detailed statistics are available for each of the trimmed and raw data files.

CGR Informatics Clinic

The CGR run an informatics clinic for our collaborators, offering one-to-one sessions lasting an hour or two, with the aim of providing limited and basic bioinformatics assistance. Please contact Sam Haldenby (s.haldenby@liverpool.ac.uk) if you would be interested in setting up a session. We also provide quoted-for, comprehensive informatics analyses, so should you require more in-depth assistance, please visit https://www.liverpool.ac.uk/genomic-research/contact-us/enquiry-form.

Description of methods

The raw Fastq files are trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 [Reference]. The option -O 3 was used, so the 3' end of any reads which match the adapter sequence for 3 bp. or more are trimmed.

The reads are further trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 15 bp. after trimming were removed. If only one of a read pair passed this filter, it is included in the R0 file. The output files from Cutadapt and Sickle are available here.

Statistics were generated using fastq-stats from EAUtils.

Additional notes

Libraries by Claudia Wierzbicki ; Sequencing by Charlotte Nelson.

Contacts

The bioinformatics analysis for this project has been performed by R.Gregory. Please e-mail if you have any further queries.