UNSW
Ramaciotti Centre
for Genomics

Hi-C Guide:

Samples requirements.

Genome assembly specifications.

UNSW SYDNEY | Australia's Global University

Contents

1 INTRODUCTION

Hi-C is a genome-wide chromatin conformation capture protocol that uses proximity ligation. It is used to generate libraries that link genomic regions that are in close spatial proximity. The sequence data from these libraries can be used to scaffold genomic assemblies, detect structural variation, and provide 3D genome conformation information.

The Ramaciotti Centre offers a Hi-C service using the Dovetail Genomics or Phase Genomics kit. Obtaining the best results from Hi-C sequence data requires both a high-quality draft genome assembly and sample input type that meets requirements. So please read this guide before harvesting your samples.

This document contains guidelines for both the Dovetail Genomics and Phase Genomics Hi-C methods. Note that recommendations are subject to change as new versions of kits are released. Please contact us to discuss your project and we can provide you with the latest guidelines and project support.

Contact us at: ramaciotti@unsw.edu.au

2 GUIDELINES FOR INPUT DRAFT ASSEMBLY QUALITY

To analyse Hi-C data, you must have a draft genomic assembly to input into the analysis pipeline. The quality of this assembly will affect the usefulness of Hi-C data. Dovetail and Phase Genomics have the following recommendations.

Dovetail Genomics Hi-C

For best results Dovetail recommends the following draft input assembly specifications when using their proprietary HiRise software:

- N50 greater than 1Mb
- N90 greater than 20kb

You can submit your Hi-C data and input draft assembly to Dovetail for analysis at cost.

Please contact the us if you would like a quote for this option.

Phase Genomics

Phase Genomics recommends the following draft input assembly for standard large eukaryotic genomes (~1Gb) specifications:

- N50 greater than 1Mb
- Draft assembly contig number below 5,000

Please see https://phasegenomics.github.io/2018/09/20/starting-assembly-guidelines.html for more information. Phase Genomics also have a site listing some of the open-source tools available: https://github.com/phasegenomics/

If you do not have a draft genome that meets these quality metrics, please contact us to discuss sequencing options for obtaining a highly quality draft genome.

3 MINIMUM COVERAGE GUIDELINES FOR GENOME ASSEMBLY

The table below provides a guide to the minimum coverage levels needed for genome assembly based on different sequencing technologies. The actual coverage needed will be highly dependent on the complexity of the genome. Please note that hybrid assembly approaches can also be performed (e.g. Illumina and PacBio sequencing).

SEQUENCING TECHNOLOGY	COVERAGE	NOTE
ONT	>50x error corrected	Recommended by Phase Genomics.
PacBIo	>70x diploid	Recommended by both Phase Genomics and Dovetail Genomics.
Illumina	High coverage >100x	Short read assemblies are acceptable in some cases.

4 SUPPORTED SPECIES

The following species and sample types are supported:

- Mammalian tissues (muscle, brain, heart, spleen)
- Mammalian cell culture
- Mammalian (non-nucleated) blood
- Non-mammalian (nucleated) blood (fish, bird, and reptile blood)
- Young plant leaves
- Insects

Please contact us if your sample type is not listed above.

See section 4 for general recommendations for each tissue. Note that each company has slightly different recommendations. Please discus your project with us before harvesting samples.

5 RECOMMENDED SAMPLE TYPES AND STORAGE CONDITIONS

(i) TISSUE

The preferred tissue types are listed below in order of most preferred to least preferred. Note this is not an exhaustive list. Please consult with Centre staff.

1. Brain
2. Spleen/ Sperm Cells
3. Kidney
4. Liver
5. Muscle (inc. Heart)

Notes

- Fat, bone or similar tissue types is not supported.
- The tissue must be taken from a live or very recently deceased specimen and snap frozen in liquid nitrogen as soon as possible after harvest.
- Frozen samples should have been stored in the -80 for less than 1 year.
- Do not use samples that have been freeze dried or preserved in RNAlater or ethanol.

Storage and Shipping

After snap freezing in liquid nitrogen, store samples at -80 degrees and ship on dry ice. Avoid freeze thaw cycles.

(ii) CELLS

- Fresh cells yield the best results.
- Avoid Freeze Thaw cycles.
- Inform the centre if using adherent cells.

Storage and Shipping

As fresh cells yield the best results it is recommended not to ship cells.

(iii) BLOOD

- Fresh whole blood samples perform better.
- Blood samples must have an anticoagulant added. EDTA is the anticoagulant of choice. Please discuss with the Centre if other anticoagulants are used.

Storage and Shipping

Fresh samples should be stored and shipped at 4 degrees and used within 2-3 days of collection. Prior arrangements will be needed to be made with The Ramaciotti Centre so that we can process the sample on arrival.

If the above is not possible, flash freeze blood samples in liquid nitrogen and keep at -80 degrees. Ship on dry ice.

(iv) PLANTS

The following is listed from most preferred to least preferred:

1. Leaves of plants at the one or two- leaf seedling stage, or cotyledons for species that are abundant in polyphenolics and/or polysaccharides, such as cotton or rose.
2. Very young leaves from more mature plants
3. Young leaves collected from plants that are pre-treated in the dark for 2-3 days.
4. Tissue from young plants other than leaves.

Notes

- Wood, bark or roots are not accepted.
- Flower buds do not produce reliable results.

Storage and Shipping

Leaves must be flash frozen in liquid nitrogen and the stored at -80 degrees. Ship on dry ice.

(v) INSECTS

The following is listed from most preferred to least preferred;

1. Embryos
2. Newly hatched larvae
3. Early pupae
4. Adults

- Data contamination from the insect's food source is a major concern. It is recommended degutting adults prior to freezing. Alternatively, they may be starved for a few days prior to freezing.
- Flash freeze individuals in liquid nitrogen, either in bulk for inbred species or individually for outbred species and store at -80 degrees. Ship on dry ice.

Storage and Shipping

Flash freeze individuals in liquid nitrogen, either in bulk for inbred species or individually for outbred species and store at -80 degrees. Ship on dry ice.

6 QUALITY ASSESSMENT

Tissues cannot be quality checked before input into the assay. Please ensure you submit samples as per guidelines.

We recommend running an iSeq sequencing run of the final Hi-C library to ensure the library is of sufficient quality before sequencing the library on a higher throughput sequencer.

We provide the data from the iSeq run to allow you to run it through the appropriate QC pipeline. We will only proceed with the final Sequencing run if you are satisfied with the results received from the QC report.

Links to QC pipelines for Dovetail and Phase Genomics are provide below.

Dovetail Genomics Omni-C QC pipeline:

https://github.com/dovetail-genomics/omni-c_qc

Phase Genomics QC pipeline:

https://phasegenomics.github.io/2019/09/19/hic-alignment-and-qc.html

7 SEQUENCING RECOMMENDATIONS

Dovetail Genomics

Dovetail recommends the following sequencing guidelines for genome scaffolding. For complex genomes or other applications, additional sequencing may be required.

	Dovetail Genomics	
Genome size (Mb)	**No. Hi-C libraries**	**No. Read pairs to sequence (Millions)**
200	1	20
500	1	50
800	1	80
1000	1	100
1500	1	150
2000	1	200
2500	2	250
3000	2	300
3500	2	350
4000	2	400
4500	3	450
5000	3	500

Phase Genomics

Phase Genomics recommends the following for human, animal, plant, fungal or microbial genome samples. For complex genomes, additional sequencing may be required.

	Phase Genomics	
Genome size (Mb)	**No. Hi-C libraries**	**No. Read pairs to sequence (Millions)**
<400	1	100
400-1500	1	150
1500-3000	1-2	250
>3000	2	Additional 100M per Gb

Notes:

- If your assembly is low contiguity (e.g., N50 <100 kb, or #contigs > 5,000), an additional 50 M read-pairs per Gb of genome size is recommended.
- For microbiome samples, Phase Genomics recommend 50 M – 100 M read-pairs for the Hi-C library and 100 M – 150 M reads for the shotgun library. If your sample is very diverse, doubling the number of both sets of reads can be useful. If your sample is of low complexity, reducing the read numbers by half is recommended.

8 SHIPPING GUIDELINES

- Only ship samples for Hi-C after pre-arrangement of project timing with Ramaciotti Staff.
- Please ship via the fastest method available (same day/overnight if available).
- Ensure all our contact details are correct and that you have forwarded on the tracking details to the contact person at the Ramaciotti Centre.
- For interstate and overseas shipping, schedule your shipments to go out at the start of the week (Monday or Tuesday).
- If you are shipping from overseas the Ramaciotti Centre can provide an import permit.
- Refer to section 5 under the specific tissue type for storage and shipping information.