

Reference Guide – Human Leukocyte Antigen (HLA) Sequencing

The human leukocyte antigen (HLA) is the human version of the major histocompatibility complex (MHC) genes found in most vertebrates.

This group of genes resides on the short arm of Chromosome 6 (see Figure 1) and codes for proteins that regulate the human immune system. Some of these proteins are denoted as antigens, due to their discovery as key factors in organ transplant tolerance.

Different classes of HLA genes are essential elements for immune function. However, only specific HLA genes corresponding to MHC Class I (HLA A, B, and C) and MHC Class II (HLA DP, DQ, and DR) present antigens that are known to have a major role in causing organ transplant rejections and providing protection mechanisms in some cancers.

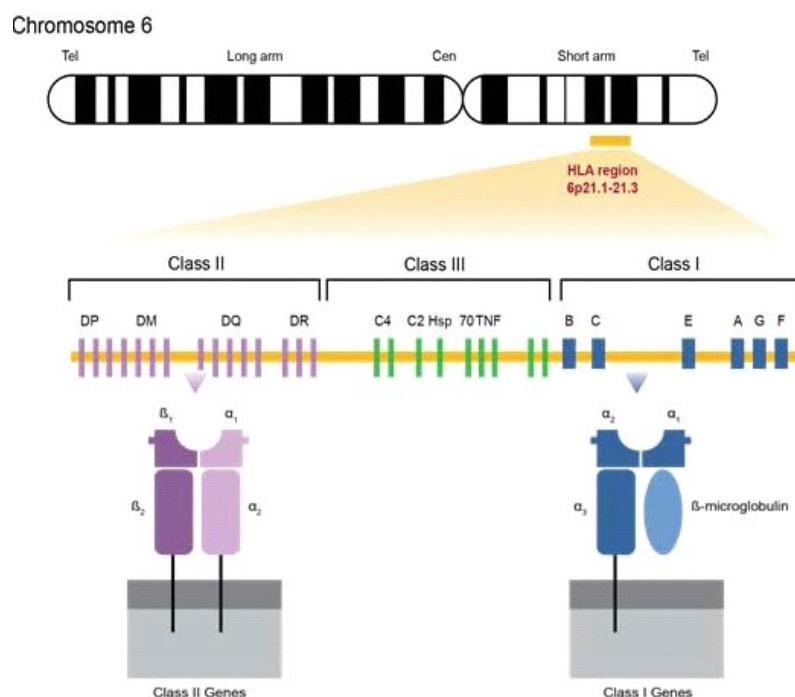


Figure 1. Structure of the HLA complex on Chromosome 6.

HLA genes are among the most polymorphic genes in the human genome. Therefore, it is very difficult to find two unrelated individuals with identical HLA molecules. As a result, HLA Class I and Class II genes are routinely sequenced or stereotyped with well-established methods for research in tissue transplantation, autoimmune disease-association studies, drug hypersensitivity research, and other applications.

This reference guide describes recommended procedures for sequencing HLA Class I and Class II genes for producing fully-phased, unambiguous, allele-level information on the Sequel®, Sequel II and Sequel IIe Systems (Sequel Systems).

HLA genes can be amplified using commercial assays (e.g., GenDx) or customers' own custom assays. Depending on the target region, full-length genes (Figure 2) or partial amplicons (Figure 3), ranging from ~1.5 kb to 11 kb, can be constructed into SMRTbell® libraries without shearing.

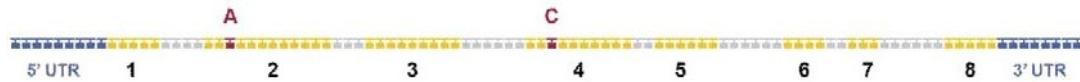


Figure 2. Fully Phased Allele-Level Sequencing of Full-length HLA Amplicons. In this example, a SNP (A) in Exon 2 is in phase with a SNP (C) in Exon 4.

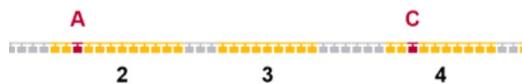


Figure 3. SNP Detection and Phasing in Exons 2, 3, and 4 Only of Partial Amplicons. In this example, a SNP (A) in Exon 2 is in phase with a SNP (C) in Exon 4.

Part A of this Reference Guide describes recommendations for GenDx HLA typing Kits. **Part B** of this Reference Guide describes recommendations for custom assays.

Part A. Recommendations for HLA Amplicons Generated with GenDx HLA Typing Kits

GenDx (gendx.com) offers validated HLA primers with ready-to-use PCR master mixes, protocols and tools for SMRT® Sequencing. Depending on your project requirements, 3 kits are available for typing up to 11 HLA loci.

GenDx NGSgo-MX6-1 Kit

PacBio® has validated the NGSgo-MX6-1 Kit, which contains primers, a ready-to-use PCR master mix and protocols. The easy-to-use kit supports the amplification of six loci (HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1) with target sizes ranging from 3 kb to 6 kb in a single tube. The 6-loci PCR product is then constructed to a SMRTbell library that can be sequenced on the Sequel Systems.

To reduce cost per sample, multiple samples may be pooled for sequencing on a single SMRT® Cell. Up to 48 barcoded samples may be pooled for sequencing on the Sequel System and up to 96 barcoded samples may be pooled for sequencing on the Sequel II and Sequel IIe Systems. See Figure 4 below for an overview of the multiplexed sequencing workflow for HLA amplicon samples.

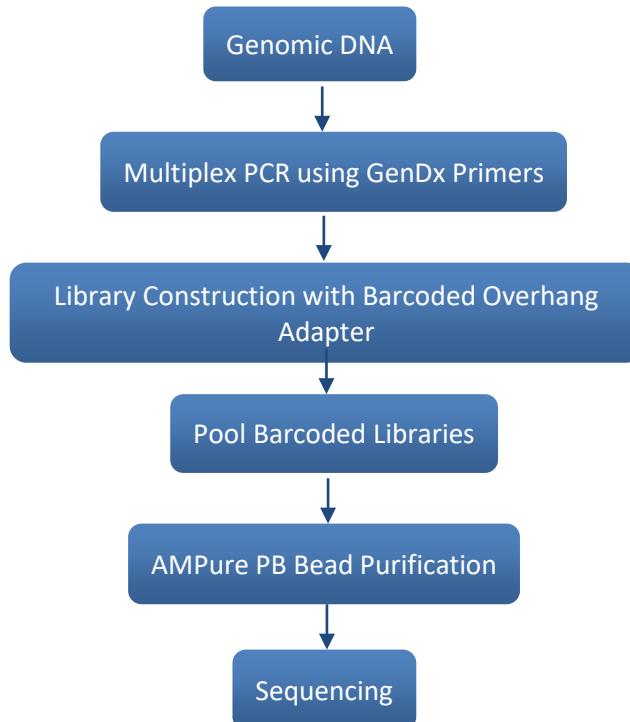


Figure 4. Workflow for amplification and sequencing of HLA amplicons amplified with GenDx primers.

Recommendations for Input Genomic DNA (gDNA)

PacBio recommends high-quality and high-molecular weight gDNA for amplification of HLA genes. Sub-optimal quality gDNA impacts not only amplification yield but also the quality of the amplified product (e.g. presence of non-specific PCR products). If not removed, non-specific PCR products can represent a substantial percentage of the sequencing reads. Below are general considerations and recommendations for gDNA samples that are to be processed for SMRT Sequencing:

- Samples have not been exposed to excessive/unnecessary freeze-thaw cycles
- Samples have not been exposed to high temperature
- Samples have not been exposed to pH extremes (<6 or >9)
- Samples do not contain insoluble material
- Samples have not been exposed to intercalating fluorescent dyes or UV radiation
- Samples do not contain chelating agents (e.g., EDTA), divalent metal cations (e.g., Mg²⁺), denaturants (e.g., guanidinium salts, phenol), or detergents (e.g., SDS, Triton-X100)
- Samples do not contain carryover contaminants from the starting organism/tissue (e.g., heme, humic acid, polyphenols)

Recommendations for Multiplexing GenDx HLA PCR Products

PacBio recommends using Barcoded Overhang Adapters (see Figure 7) for multiplex sequencing on the Sequel Systems. SMRTbell adapters with 16-bp barcode sequences are ligated to both ends of PCR products during library construction to enable sample pooling, and allow sequencing reads to be identified and sorted for data analysis.

Up to **48 barcoded samples** may be pooled for sequencing on the Sequel System and **up to 96 barcoded samples** may be pooled for sequencing on the Sequel II and Sequel IIe Systems.

Recommendations for Barcoded Overhang Adapters

- **Barcoding Up to 16 Samples (with Barcoded Overhang Adapter Kits 8A and 8B)**

PacBio offers two commercially available barcoded overhang adapter kits: 1) Barcoded Overhang Adapter Kit 8A and 2) Barcoded Overhang Adapter Kit 8B. Each kit includes 8 ready-to-use, validated Barcoded Overhang Adapters. The workflow for barcoding amplicons samples for multiplexed SMRT Sequencing is described in [Procedure & Checklist - Preparing SMRTbell Libraries using PacBio Barcoded Overhang Adapters for Multiplexing Amplicons](#).

- **Barcoding >16 Samples**

If your project requires multiplexing more than 16 samples, there are available 96 barcoded adapter sequences that can be ordered for synthesis from your preferred Oligo Synthesis Provider (see our [Multiplexing Resources](#) website). The list contains sequences (5'→3') of barcoded overhang adapter sequences along with recommended synthesis scale and oligo purification methods.

Before using, the synthesized oligos must be diluted to a working stock and annealed to form hairpin adapters. The annealing procedure is outlined in [Procedure & Checklist - Preparing SMRTbell Libraries using PacBio Barcoded Overhang Adapters for Multiplexing Amplicons](#).

SMRTbell Library Construction

To proceed to library construction, you must have the following:

- Downloaded and reviewed the [Procedure & Checklist - Preparing SMRTbell Libraries using PacBio Barcoded Overhang Adapters for Multiplexing Amplicons](#)
- Barcoded Overhang Adapters
- 150 ng of PCR product per sample to be barcoded
- SMRTbell Express Template Prep Kit 2.0

1. The SMRTbell library construction procedure starts on page 5 “Concentration of PCR Products by AMPure® PB Beads”. Purification of PCR products with AMPure PB beads is highly recommended so that samples can be pooled in equimolar concentrations for SMRTbell library construction. Note that **150 ng of PCR product** per sample (comprised of a single amplicon target or pool of amplicon targets) is required in order to proceed with the first enzymatic step (DNA Damage Repair).
2. DNA Damage Repair and End-Repair steps are described on page 8.
3. The Adapter Ligation step with Barcoded Overhang Adapters is described on page 9.
4. After ligation, the barcoded SMRTbell libraries are pooled as described on page 9.
5. The pooled barcoded library is then purified using AMPure PB beads described on pages 10 to 11. Two AMPure PB bead purifications are required after pooling.
 - a. For the first-round of AMPure purification, use 400 µL (instead of 100 µL) Elution Buffer to elute the DNA more efficiently.
 - b. For the second-round of AMPure purification, use 100 µL (instead of 50 µL).
6. Finally, the purified library is ready for sequencing. For sample setup and run design recommendations, refer to the next section.

DNA Sequencing Recommendations for the Sequel Systems

Instructions for preparing SMRTbell libraries for sequencing are found in SMRT® Link v9.0 or higher. Specifically, SMRT Link Sample Setup provides guidance on how to anneal the sequencing primer, bind the sequencing polymerase and clean up the sample for loading on the PacBio instrument.

- **SMRT Link Sample Setup Conditions for Primer Annealing and Polymerase Binding**

SMRT Link Version	Sample Setup Parameter	Sequel System	Sequel II and Ile Systems
v10.0	Application	N/A	Select ‘≥3 kb Amplicons’
v9.0 / v10.0	Sequencing Primer	Sequencing Primer v4	Sequencing Primer v4
	Primer:Template Ratio	20:1	20:1
	Binding Kit	Sequel Binding Kit 3.0	Sequel II Binding Kit 2.0
	Polymerase:Template Ratio	30:1	30:1
	Loading concentration (pM)	5	50-60

- **SMRT Link Run Design Settings and Recommended Sequencing Run Parameters**

SMRT Link Version	Run Design	Sequel System	Sequel II and Ile Systems
v10.0	Application	N/A	Select ‘≥3 kb Amplicons’
	Generate HiFi Reads (SMRT Link v10.0)	N/A	Select ‘On-Instrument’ or ‘In SMRT Link’
v9.0	Sequencing Mode (SMRT Link v9.0) Generate CCS Data (SMRT Link v9.0)	Select ‘CCS Reads’ Select ‘Yes’ to enable auto-CCS analysis	Select ‘CCS Reads’ Select ‘Yes’ to enable auto-CCS analysis
	Pre-extension time (hrs)	1.1	1.1
v9.0 / v10.0	Movie time (hrs)	20	20

Data Analysis and NGSengine Typing Software

PacBio recommends GenDx's NGSengine software (https://www.gendx.com/product_line/NGSengine/) for HLA typing. With SMRT® Analysis and NGSengine, data analysis and typing are simple and easy.

Demultiplexing of barcodes is automatically performed by SMRT Analysis (v8.0 or higher). Highly accurate long reads (HiFi) are generated by selecting reads that meet the $\geq Q20$ filtering criteria. The resulting HiFi FASTA/Q files are then imported to NGSengine for typing.

Below is a summary of the analysis workflow:

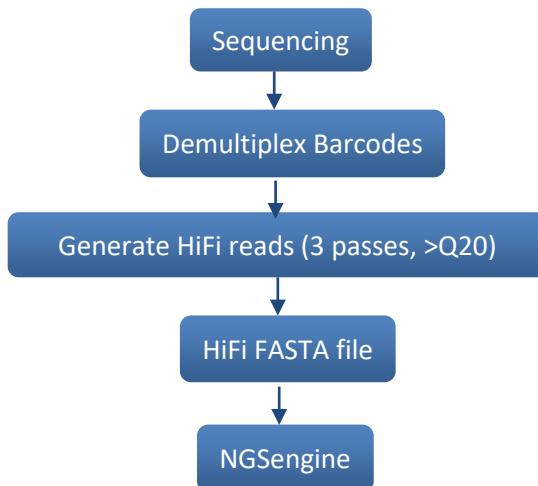


Figure 5. High level analysis workflow for HLA typing.

GenDx's NGSengine allows mapping of each HLA locus, calls SNPs, provides coverage statistics and generates an 8-digit level of typing with high accuracy and confidence. The typing software also generates comprehensive reports for data release. Recommended settings for analysis of PacBio HiFi data are shown below in Figure 6.

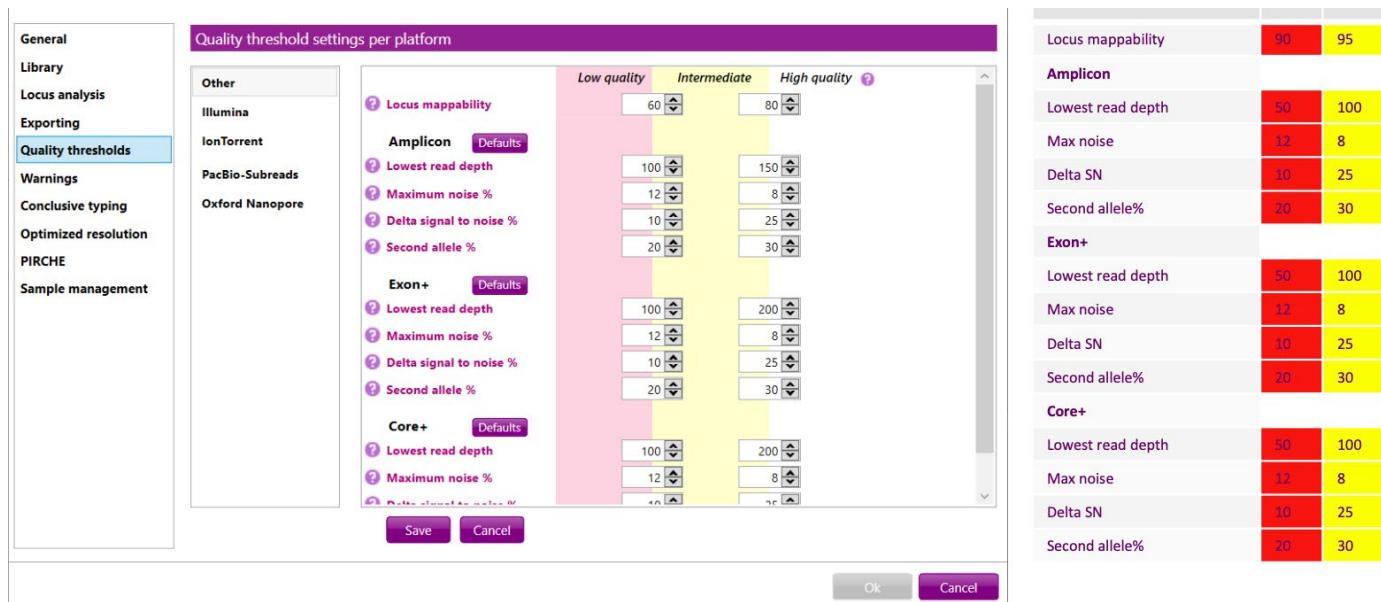


Figure 6. Recommended PacBio CCS quality threshold.

Part B. Recommendations for HLA Amplicons Generated with Customer-Designed Assays

Note: Users may design their own custom primers to amplify HLA gene targets for SMRT Sequencing – but be aware that painstaking and time-consuming validation experiments may be required to develop an optimized sample preparation workflow. We highly recommend using validated and ready-to-use primers from GenDx instead to amplify HLA gene targets for PacBio sequencing (See **Part A** of this Reference Guide).

Any customer-designed HLA assays are compatible with SMRT Sequencing. PCR products comprised of HLA amplicons can be constructed to SMRTbell libraries and sequenced on the Sequel or Sequel II and Ile Systems.

Cost per sample may be reduced by incorporating barcodes into each sample, allowing pooling for sequencing on a single SMRT Cell (8M). Three barcoding strategies along with their respective procedures are described below:

1. Barcoded Overhang Adapters ([Procedure & Checklist](#)): This procedure is recommended for HLA amplicon samples generated using commercial GenDx primers or any PCR products generated from customer-validated custom primers. Samples are ligated with barcoded adapters and then pooled to create a multiplexed SMRTbell library for sequencing.
2. Barcoded Universal Primers ([Procedure & Checklist](#)): This procedure enables symmetric barcoding of up to 96 amplicon samples using PacBio Barcoded Universal Primers.
3. Barcoded M13 Primers ([Procedure & Checklist](#)): This procedure enables asymmetric barcoding of up to 1024 amplicon samples using a combination of 32 forward and 32 reverse Barcoded M13 Primers.

All three workflows listed above include barcodes to uniquely identify samples within a pooled library mixture, however, they differ in how the barcodes are incorporated into the SMRTbell template.

Barcoded Overhang Adapter Workflow

Barcodes are introduced to each sample through ligation during SMRTbell library construction (see Figure 7). The resulting barcoded sample may be pooled with other samples for sequencing on the Sequel or Sequel II and Ile Systems.

There are currently 16 Barcoded Overhang Adapters that are commercially available from PacBio:

- PacBio Barcoded Overhang Adapter Kit 8A and
- PacBio Barcoded Overhang Adapter Kit 8B

If multiplexing projects require more than 16 samples, we offer a list of 96 barcoded overhang adapter sequences that can be ordered from an oligo synthesis provider (see our [Multiplexing Resources](#) website). The list contains sequences (5'→3') of barcoded overhang adapters along with their recommended synthesis scale and purification method.

The annealing procedure for forming SMRTbell hairpins is also available in the “Annealing PacBio SMRTbell Overhang Adapters” section of [Procedure & Checklist - Preparing SMRTbell Libraries using PacBio Barcoded Overhang Adapters for Multiplexing Amplicons](#).

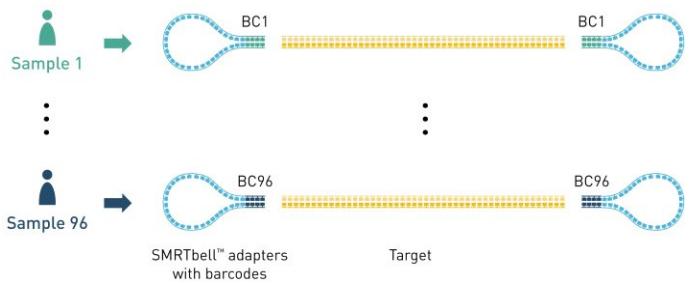


Figure 7. Barcoding using Barcoded Overhang Adapters. Incorporation of barcodes takes place during ligation in the SMRTbell library construction process.

Barcoded Universal Primer Workflow

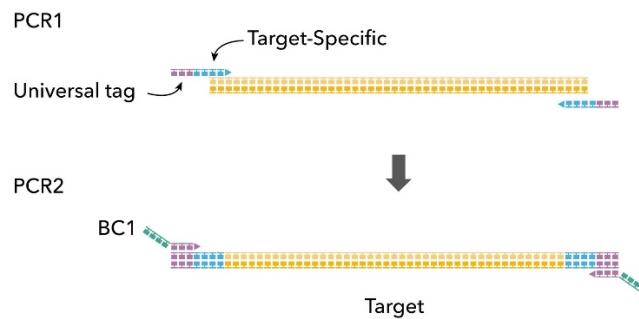
Barcodes are incorporated into the amplicon via a two-step tailed PCR approach. This workflow uses a symmetric barcoding strategy which integrates the same barcode sequence to both ends of the SMRTbell template.

As shown in Figure 8, internal primers that are a combination of universal (30 bp) and target-specific sequences are used for the first round of PCR (PCR1). After the universal sequences are incorporated into the PCR amplicon during PCR1, a second round of PCR (PCR2) is performed which uses barcoded universal primers to incorporate the barcodes. A Barcoded Universal Primers kit containing 96 barcoded universal primers can be purchased from PacBio. The part number of the kit is available in the [Barcoded Universal Primers Procedure & Checklist](#).

Barcoded M13 Primer Workflow

The Barcoded M13 Primer workflow is similar to the Barcoded Universal Primer workflow, except that it uses an asymmetric barcoding strategy (different barcode sequences are incorporated on each side of the template). Additionally, the “universal” sequence in this case is a 17 bp sequence called M13. Target-specific primers used in the first-round PCR step are tailed with the M13 sequence and can be custom-ordered from your preferred oligo synthesis provider. The list of M13 primers tailed with PacBio barcodes for the second-round PCR step is available in the [Multiplexing Resources](#). By using a combination of 32 forward and 32 reverse Barcoded M13 Primers, up to 1024 barcoded HLA samples may be pooled for sequencing on a single SMRT Cell 8M.

Sample 1



Sample N

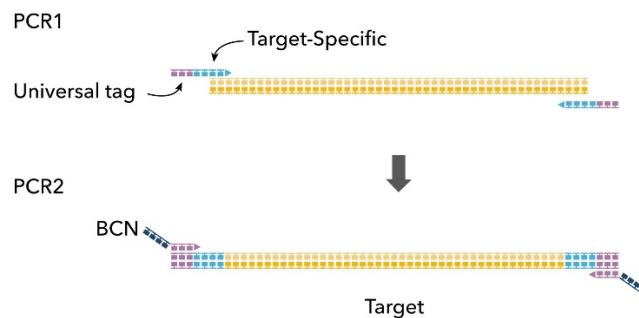


Figure 8. Barcoding using Barcoded Universal Primers (or Barcoded M13 Primers) in a 2-step PCR method. Incorporation of barcodes takes place during PCR2 (second-round PCR).

Recommendations for Input Genomic DNA (gDNA) QC

PacBio recommends high quality and high molecular weight gDNA for amplification of HLA genes. Sub-optimal quality gDNA impacts not only amplification yield but also the quality of the amplified product (e.g. presence of non-specific PCR products). Below are general considerations and recommendations for gDNA samples that are to be processed for SMRT Sequencing:

- Samples have not been exposed to excessive/unnecessary freeze-thaw cycles.
- Samples have not been exposed to high temperature
- Samples have not been exposed to pH extremes (<6 or >9)
- Samples do not contain insoluble material
- Samples have not been exposed to intercalating fluorescent dyes or UV radiation
- Samples do not contain chelating agents (e.g., EDTA), divalent metal cations (e.g., Mg²⁺), denaturants (e.g., guanidinium salts, phenol), or detergents (e.g., SDS, Triton-X100)
- Samples do not contain carryover contaminants from the starting organism/tissue (e.g., heme, humic acid, polyphenols)

Recommendations for PCR Amplification

Clean, target-specific PCR products are extremely important for obtaining high-quality sequence data. Non-specific products can represent a substantial percentage of the sequencing reads if they are not removed. To minimize the presence of non-specific products, consider the following recommendations for generating high-quality amplicons suitable for SMRTbell library preparation and sequencing.

1. For PCR amplification, KAPA HiFi HotStart 2x ReadyMix PCR Reagent is recommended. Thaw on ice and mix well before use. Note: All KAPA HiFi HotStart reagents and reactions must be set up and kept on ice until PCR; the high proofreading activity of the enzyme will rapidly degrade primers at room temperature.
2. Begin with high-quality nucleic acids and work in a clean environment.
 - a. If extracted nucleic acids must be stored, freeze at high concentrations in appropriately-buffered solutions.
 - b. To minimize possible contamination and degradation caused by multiple freeze/thaw cycles, sub-aliquot DNA into smaller volumes for storage.
 - c. Set up PCR reactions in an environment free from sources of non-specific primer and template contaminants; ideally a laminar flow hood, using dedicated pre-PCR pipettor, tips and reagents.
3. Use PCR reagents and conditions for generating target-specific, full-length amplicons.
 - a. Use the highest-fidelity polymerase compatible with your PCR amplification system.
 - b. Use desalted or HPLC-purified oligos; damaged bases at the ends of the amplicons cannot be repaired by DNA Damage Repair enzymes.
 - c. Optimize PCR conditions to minimize total time spent at high (>65°C) temperatures, particularly during denaturation.
 - d. PCR extension time should be long enough to ensure complete extension, taking into consideration the polymerase used and target amplicon size. For mixed samples with similar targets, it is important to complete extension at every step to avoid generating chimeric products in subsequent steps. As a general guideline, use extension times of one minute per 1000 base pairs (e.g., 3 minutes for a 3 kb product).
4. Use the lowest number of cycles required for obtaining adequate yields (ng) of PCR products to proceed with SMRTbell library construction. Avoid over-amplification.
5. If non-specific products are present, optimize PCR conditions or perform AMPure PB Bead-based size selection to enrich for PCR amplicons with the desired target size (see recommendations below).

Amplicon Pooling Recommendations for SMRTbell Library Construction

Option 1: Multiplexing Amplicons of the Same Size

Pool amplicons of similar size, generally within 15% the size of each other. For example, Class I genes (HLA-A, HLA-B and HLA-C) are approximately 3.2 to 3.5 kb in size and may be pooled for SMRT Sequencing. See Figure 9 below.

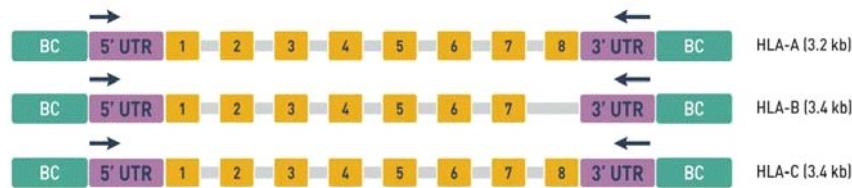


Figure 9. Pooling Class I Genes. Class I genes have similar amplicon sizes that can be multiplexed and sequenced on a single SMRT Cell to generate full-length HLA genes (~3.2 to 3.5 kb HLA Class I genes).

It is a best practice to measure the DNA concentration of each sample to facilitate equimolar pooling for SMRTbell library construction. Ideally, amplicon samples should be AMPure PB bead purified and quantified prior to pooling.

Alternatively, PCR product yield and quality may be visually assessed using agarose gel electrophoresis (see Figure 10). This QC method may be employed if the amplicon targets have the same size (+/- 15%) and have similar amplification efficiencies, allowing pooling by equal volume. For a sample showing a weak signal on a gel, increasing the volume of the sample for pooling can help reduce biased representation in the sequencing data.

If the PCR yield across all samples appears consistent on a gel, samples may be pooled by equal-volume.

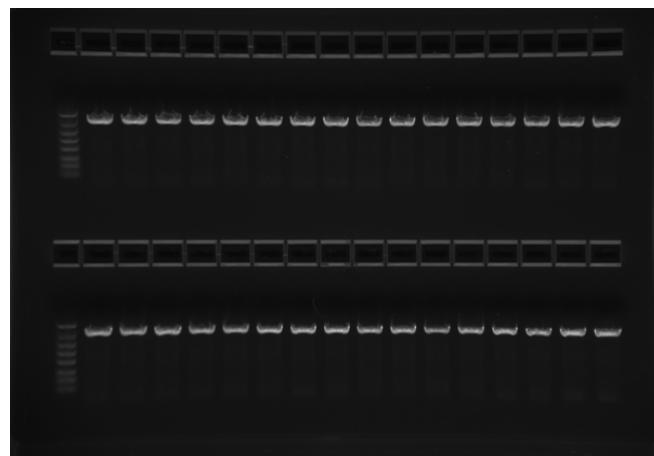


Figure 10. HLA A amplicons loaded on an agarose gel (1.2% Lonza FlashGel) for visual inspection of PCR product yields.

Option II: Multiplexing Mixed-Size Amplicons (Class I and Class II)

Class I (3 kb) and Class II (up to 8 kb), may also be pooled for sequencing on one SMRT Cell (See Figure 11). For best results and even coverage, it is recommended to pool amplicons at equimolar concentrations for SMRTbell library construction. Ideally, samples should be AMPure PB bead purified and quantified to facilitate equimolar pooling.

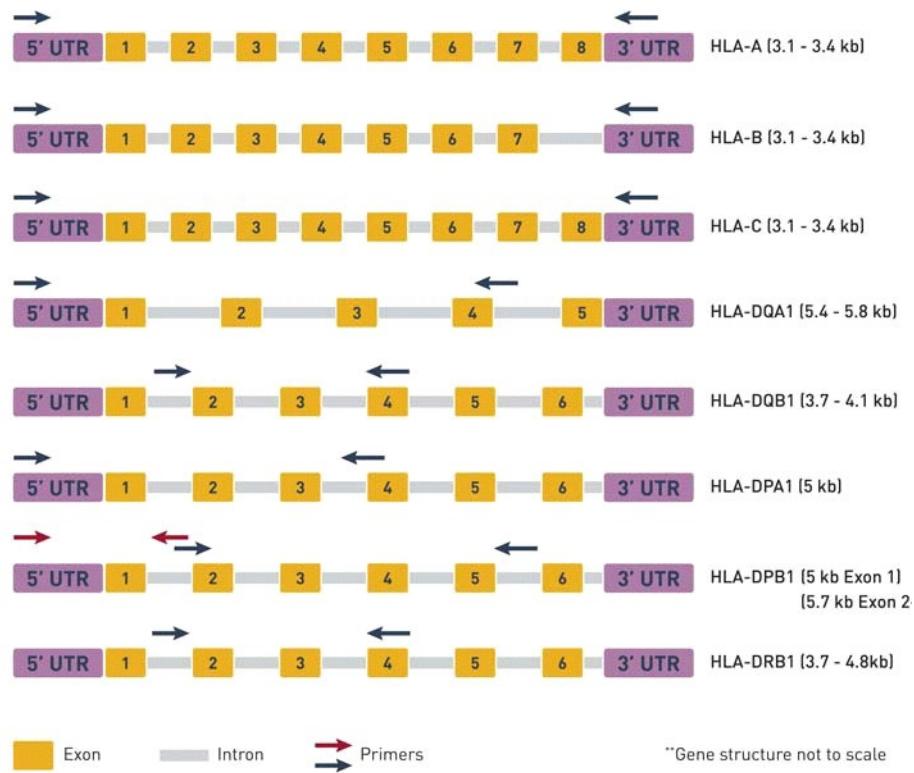


Figure 11. Example of Class I and II genes with varying lengths. Pool samples in equimolar concentrations.

High-Throughput Workflow for Multiplexed HLA Sequencing

As your HLA sequencing throughput needs increase, the sample multiplexing level may be further increased by using the workflow illustrated in Figure 12.

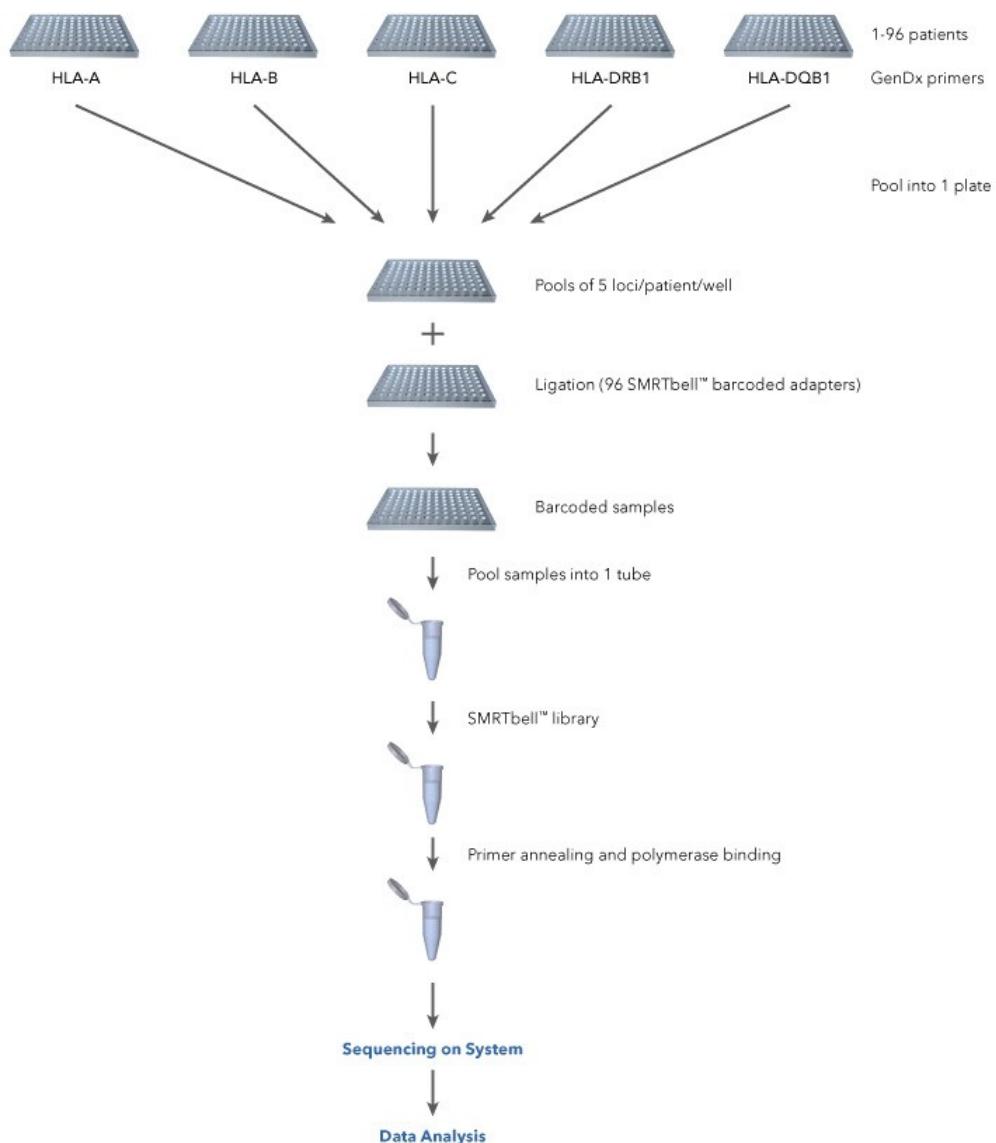


Figure 12: In this example, 96 DNA samples are amplified using each primer set (HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1). After amplification, PCR products from the five plates (A, B, C, DRB1, DQB1) are pooled – Ideally in an equal-molar fashion – into one 96-well plate (with each well containing 5 loci per sample. For SMRTbell library construction, each well is individually processed and barcoded using barcoded adapters. Finally, the 96 uniquely-barcoded samples are pooled into a single tube for sequencing on the Sequel Systems.

DNA Sequencing Recommendations for the Sequel and Sequel II/ Ile Systems

Instructions for preparing SMRTbell libraries for sequencing are found in SMRT Link v9.0 or higher. Specifically, SMRT Link Sample Setup provides guidance on how to anneal the sequencing primer, bind the sequencing polymerase and clean up the sample for loading on the PacBio instrument.

- **SMRT Link Sample Setup Conditions for Primer Annealing and Polymerase Binding**

SMRT Link Version	Sample Setup Parameter	Sequel System	Sequel II and Ile Systems
v10.0	Application	N/A	Select '≥3 kb Amplicons'
v9.0 / v10.0	Sequencing Primer	Sequencing Primer v4	Sequencing Primer v4
	Primer:Template Ratio	20:1	20:1
	Binding Kit	Sequel Binding Kit 3.0	Sequel II Binding Kit 2.0
	Polymerase:Template Ratio	30:1	30:1
	Loading concentration (pM)	5	50-60

- **SMRT Link Run Design Settings and Recommended Sequencing Run Parameters**

SMRT Link Version	Run Design	Sequel System	Sequel II and Ile Systems
v10.0	Application	N/A	Select '≥3 kb Amplicons'
	Generate HiFi Reads (SMRT Link v10.0)	N/A	Select 'On-Instrument' or 'In SMRT Link'
v9.0	Sequencing Mode (SMRT Link v9.0) Generate CCS Data (SMRT Link v9.0)	Select 'CCS Reads' Select 'Yes' to enable auto-CCS analysis	Select 'CCS Reads' Select 'Yes' to enable auto-CCS analysis
v9.0 / v10.0	Pre-extension time (hrs)	1.1	1.1
	Movie time (hrs)	20	20

Data Analysis and NGSEngine Typing Software

PacBio recommends using GenDx's NGSEngine software (https://www.gendx.com/product_line/NGSEngine/) for HLA typing. With SMRT® Analysis and NGSEngine, data analysis and typing are simple and easy. Demultiplexing of barcodes is automatically performed by SMRT Analysis (v8.0 or higher). Highly accurate long reads (HiFi) are generated by selecting reads that meet the ≥Q20 filtering criteria. The resulting HiFi FASTA/Q files are then imported to NGSEngine for typing.

Below is a summary of the analysis workflow:

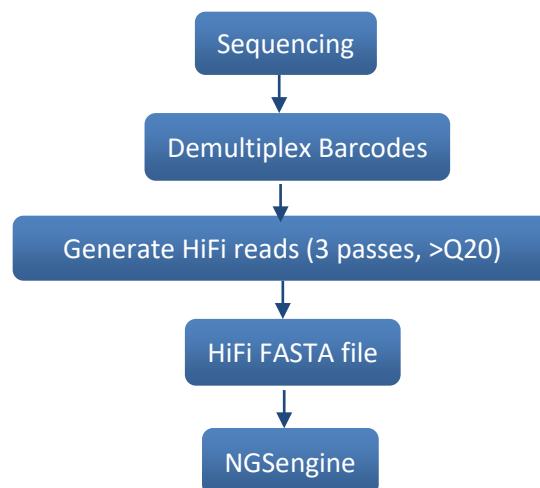


Figure 13. High level analysis workflow for HLA typing.

GenDx's NGSEngine allows mapping of each HLA locus, calls SNPs, provides coverage statistics and generates an 8-digit level of typing with high accuracy and confidence. The typing software also generates comprehensive reports for data release. Recommended settings for analysis of PacBio HiFi data are shown below in Figure 14.

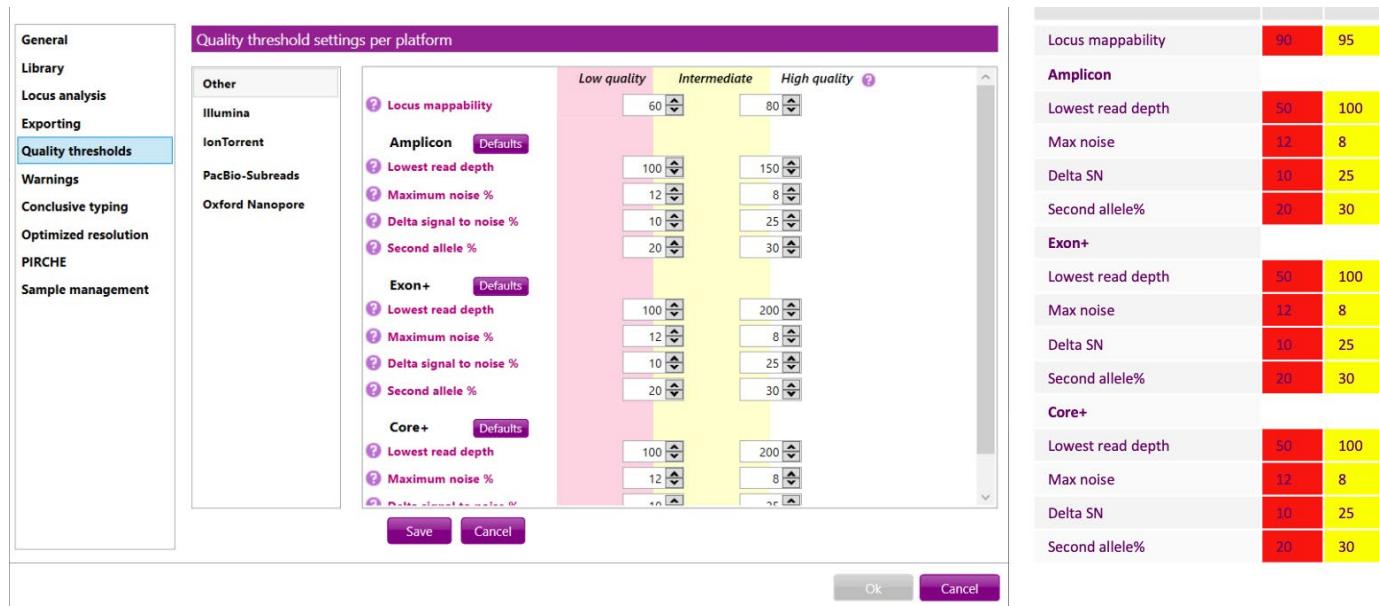


Figure 14. Recommended PacBio CCS quality threshold.

References

PacBio Web Page: <https://www.pacb.com/applications/targeted-sequencing/hla/>

Mayor, N. P., et al. (2019). [Recipients receiving better HLA-matched hematopoietic cell transplantation grafts, uncovered by a novel HLA typing method, have superior survival: A retrospective study.](#) Biology of Blood and Marrow Transplantation, 25(3), 443–450.

Webinar: Understanding SARS-CoV-2 and host immune response to COVID-19 with PacBio sequencing, <https://youtu.be/Unm3Rg7fPYg>

Poster: [McLaughlin, I., et al. \(2017\) Targeted sequencing using a long-read sequencing technology](#)

Revision History (Description)	Version	Date
Initial release.	01	January 2021

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2021, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at <https://www.pacb.com/legal-and-trademarks/terms-and-conditions-of-sale/>. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq and Sequel are trademarks of Pacific Biosciences. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.