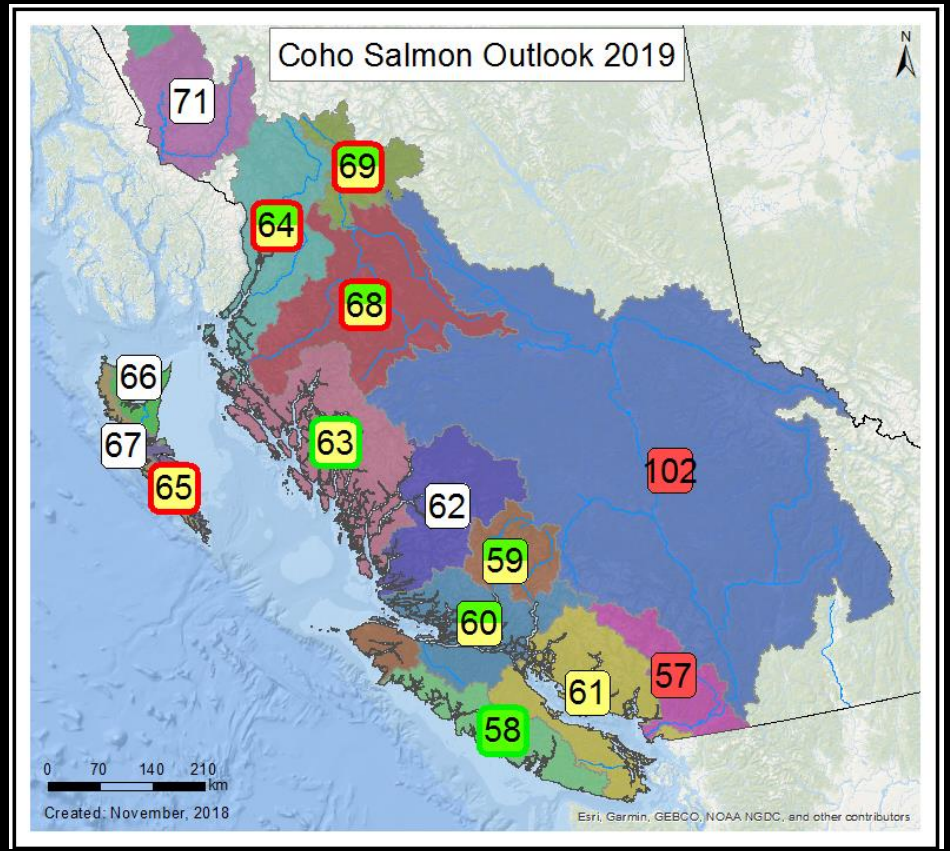
At sea genetic stock identification of overwintering coho salmon in the Gulf of Alaska: Evaluation of nanopore sequencing for remote real-time deployment

PICES W16 Oct 22nd

Christoph Deeg, Ben Sutherland, Tobi Ming, Collin Wallace, Kim Jonsen, Kelsey Flynn, Charlie Waters, Richard Beamish, Terry Beacham, and Kristi Miller

Stock Identification

- Needed for management of:
  - Harvest
  - Enhancement
  - Conservation
- Methods:
  - Scales, parasites, CWT
  - Allozymes
  - Genetic
    - Mini- / Micro-satellites
    - SNPs -> Highest resolution and accuracy



Coho conservation units

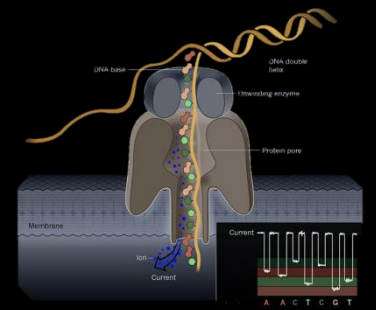
GSI by SNP sequencing

- Genetic stock identification (GSI) by Single Nucleotide Polymorphism (SNP) sequencing
  1. Sequence (e.g. RAD-Seq) of representative populations
  2. Identify SNPs
  3. Primer pool to amplify SNPs
  4. Sequence SNPs of individuals from known origin (-> baseline)
  5. Sequence SNPs of individuals in question (-> query)
  6. Compare query with baseline -> assign to stock

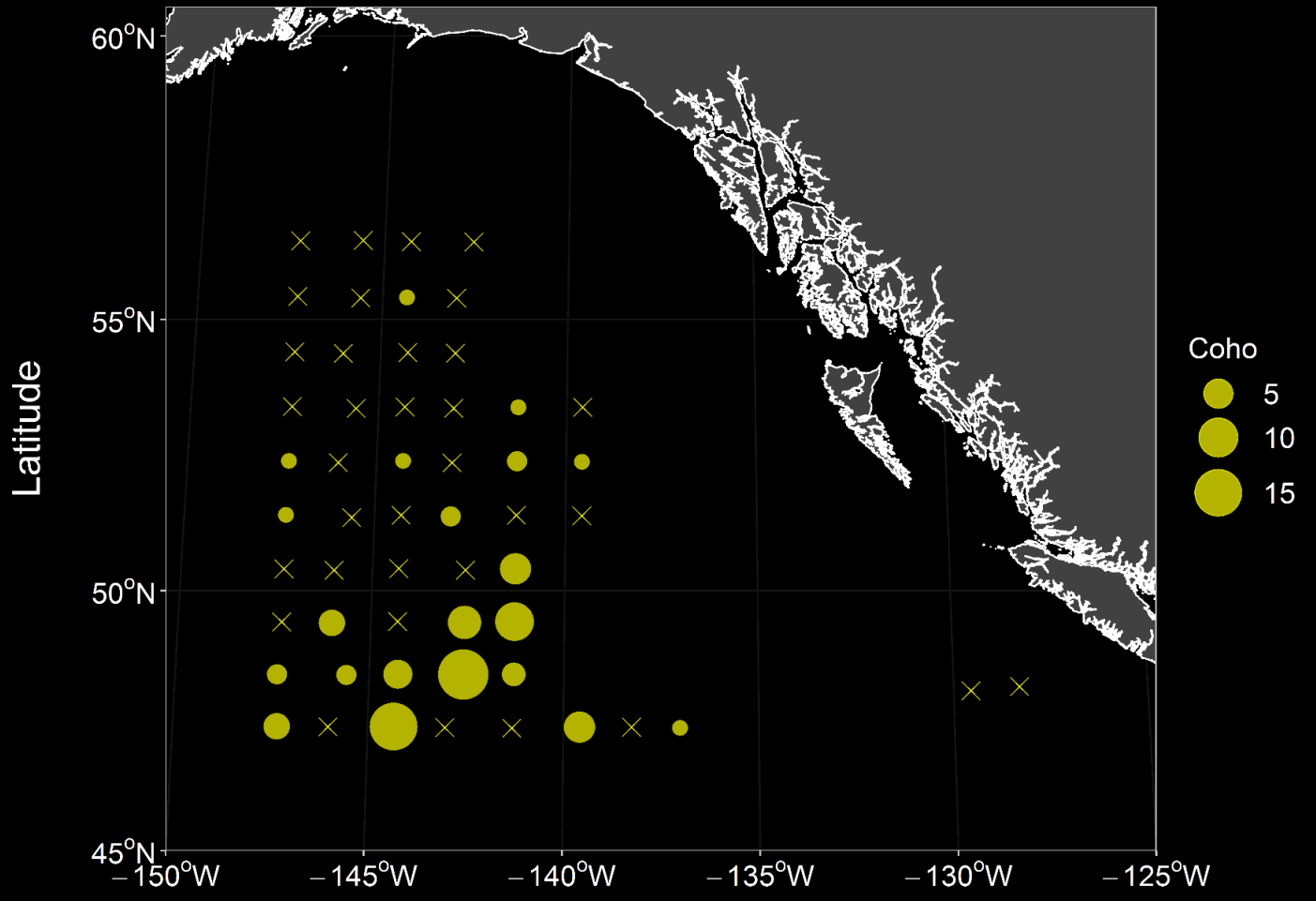
Why use nanopore for GSI?

- Ion Torrent (semiconductor) based SNP GSI
  - Genotyping by the thousands (GT-Seq) -> high throughput
  - Disadvantage:
    - Large batches
    - Complex infrastructure
- Goal: In-field “real-time” SNP GSI
  - Fast and flexible genotyping with minimal infrastructure
- Nanopore sequencing
  - Disadvantages for GSI
    - High error rate
    - Limited pores available vs. Short amplicons

  => concatenate amplicons!

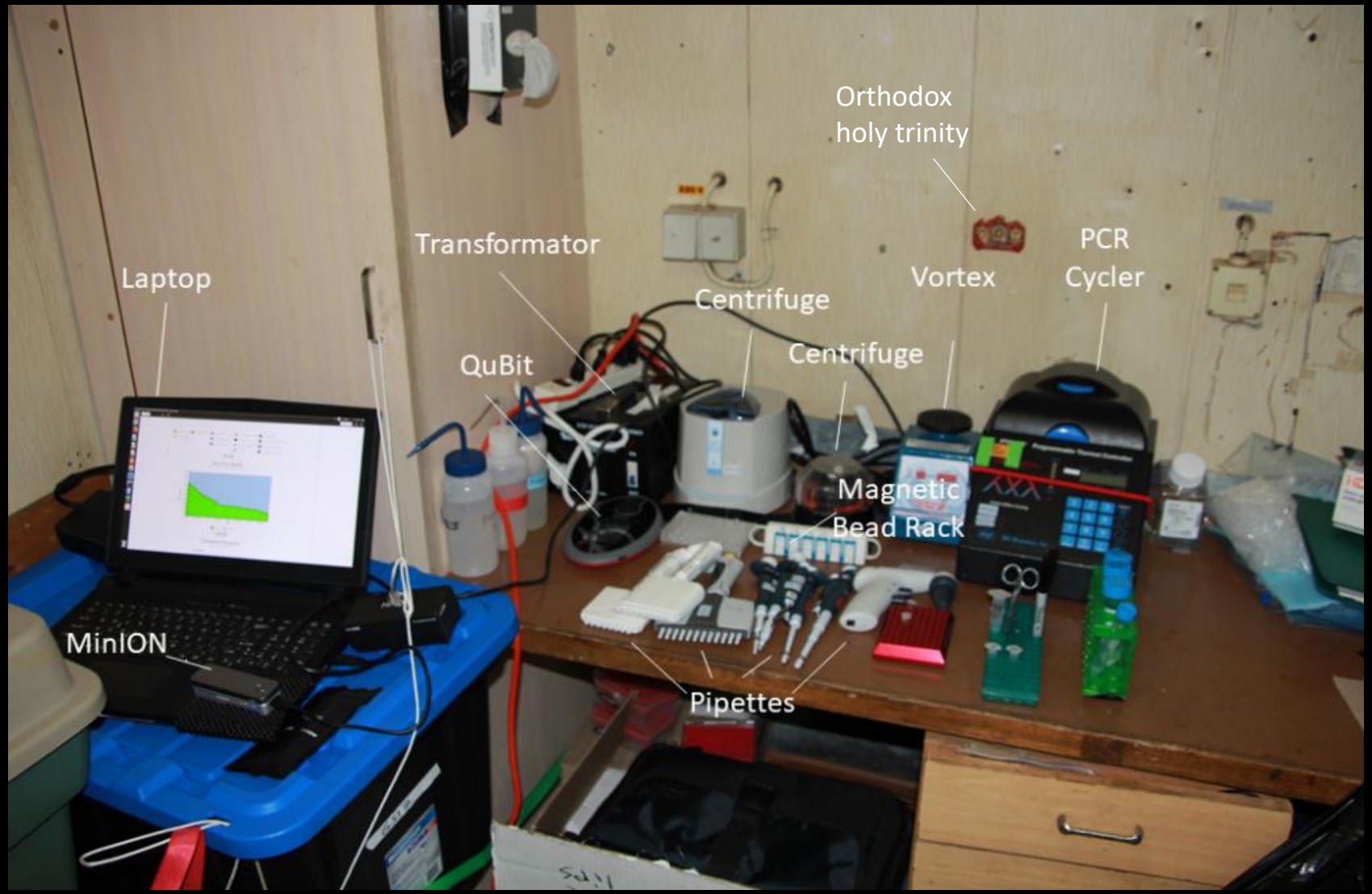


Coho caught



Charlie Waters

GSI setup


Orthodox
holy trinity
Transformator
Laptop
Vortex
PCR
Cycler
Centrifuge
QuBit
Centrifuge
Magnetic
Bead Rack
MinION
Pipettes

Wetlab

DNA extraction

SNP

PCR amplification

Other fish

Barcoding (Fish ID) and pooling

Inverse adapter ligation

\+

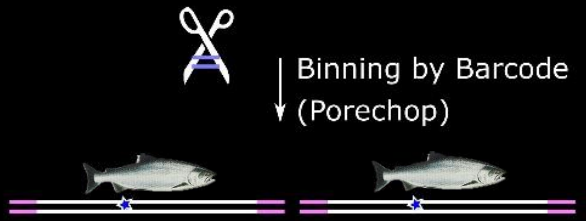
Hetero-dimer priming

Concatenation

Sequencing

Computational analysis

Deconcatenation (Porechop)



Binning by Barcode (Porechop)

Align to reference (bwa)

SNP

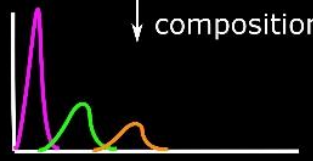
Score SNPs (n2g)

0/1;1/1/;0/0

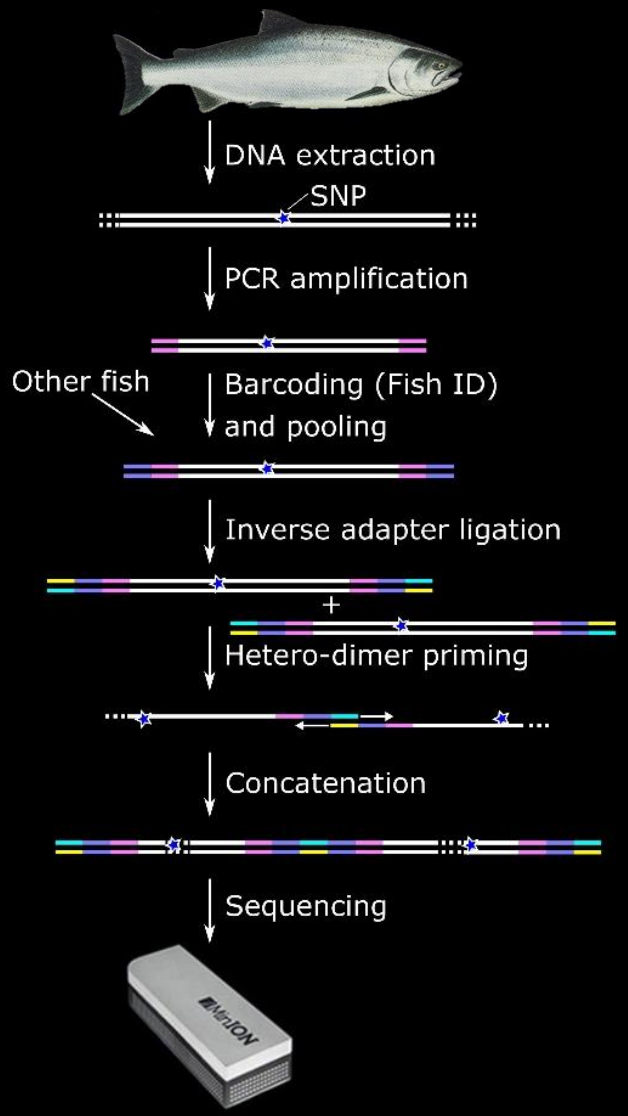
Assign to referencr populations (rubias)

0/1;1/1/;0/0
1/1;1/1/;0/1

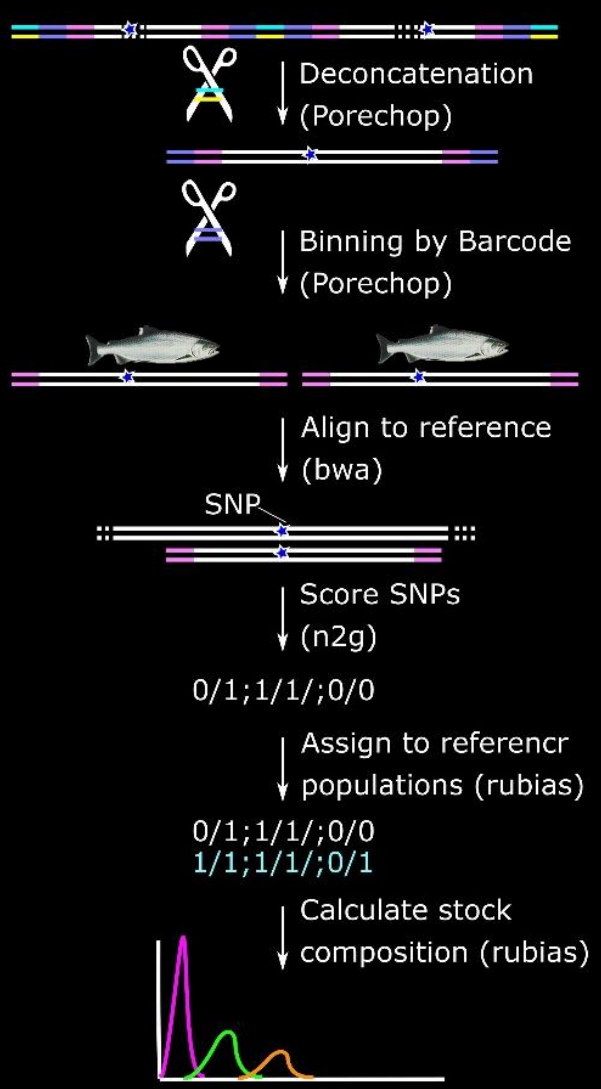
Calculate stock composition (rubias)



Wetlab Workflow



- Extract DNA
  - Quickextract (20min)
- Multiplex PCR
  - V3 coho (296), V6 chinook (299))
  - AgriSeq kit
- Barcoding (fish-ID)
  - ONT 96 Barcoding kit -> Blunt end ligation
- Blunt end ligation of inverse adapters
- PCR-like concatenation
- Nanopore adapter ligation
- MinION Sequencing



Computational Workflow (nano2geno)

- Basecalling with MinKNOW
  - Ubuntu 14.06, 31.2 GiB RAM 7700K CPU @ 4.20GHz × 8
- Deconcatenate and bin
  - Porechop
    - Custom barcode files
    - Default binning
- Align with BWA
- Score with custom script n2g (Ben Sutherland)
- Assign to stock with rubias

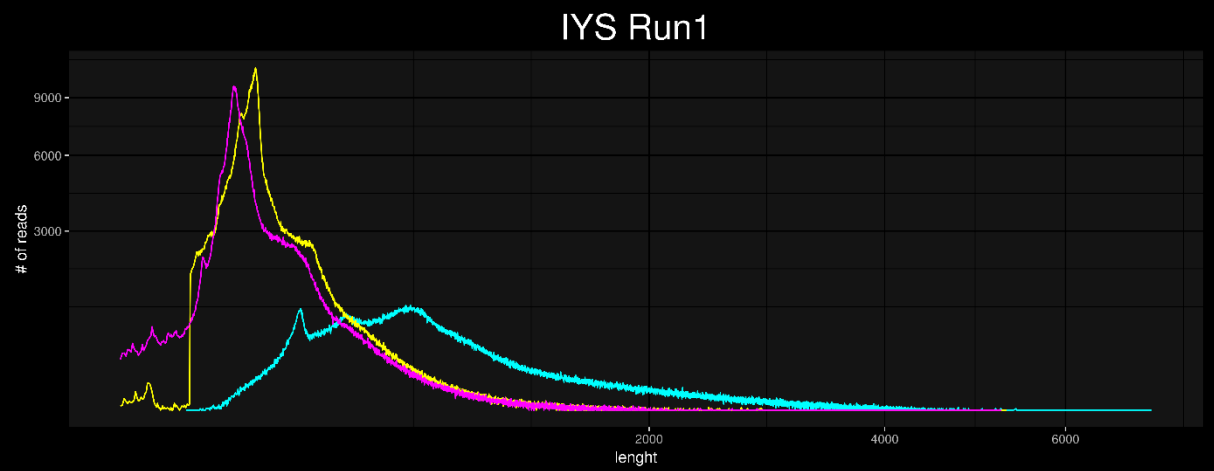
GSI runs analyzed

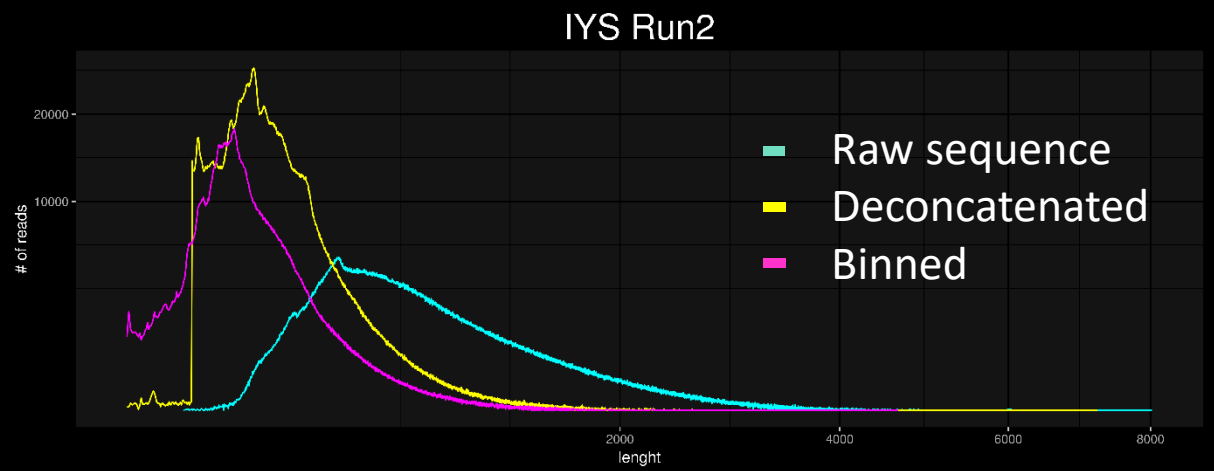
- 1st IYS nanopore run: 31 coho + 2 Chinook
- 2nd IYS nanopore run: 44 coho + 1 Chinook
- PBS nanopore: Resequenced all 80 coho
- Resequencing on IonTorrent with matching baseline / primers
- Independent sequencing of all IYS coho with new primer set and baseline

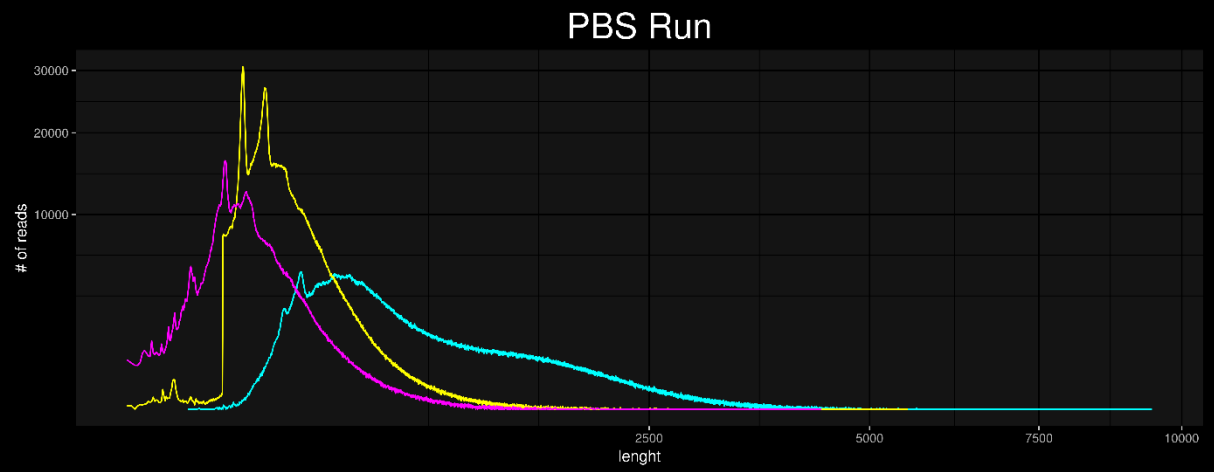
Faulty flow cell priming

Results: Reads and Concatenation

- 1.5 – 5.38M reads
- 1.5 – 2 x inflation after deconcatenation
- Inefficient Barcoding!
  - None bin 12-50%!

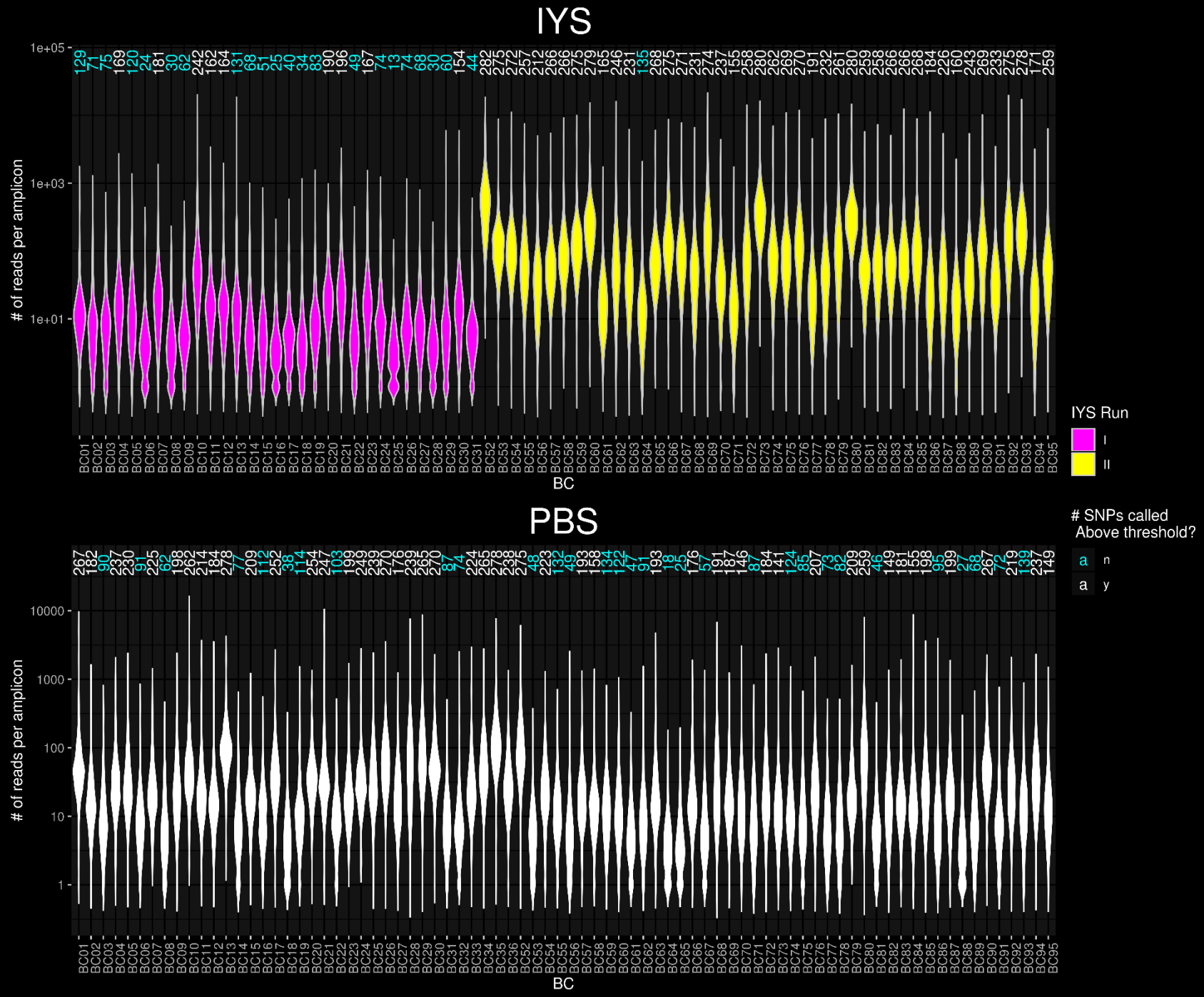






Read distribution

- Read distribution somewhat even
  - No normalization!
- Need > 2000 reads for >50% (141) of loci at 10x
  - IYS1: 9/31 (29%)
  - IYS2: 43/44 (98%)
  - PBS: 50/80 (63%)

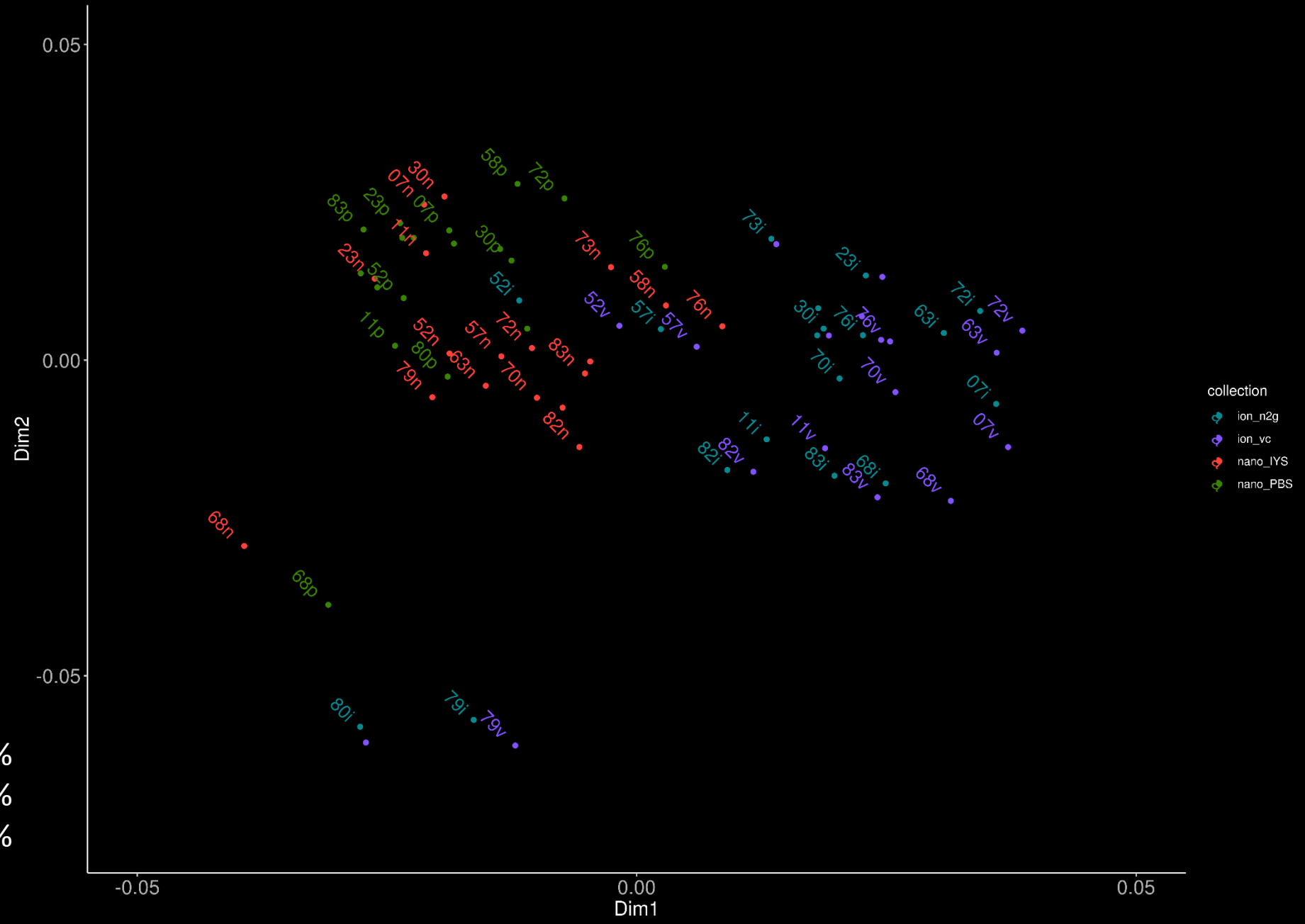


Problematic loci

Homopolymer containing loci and loci problematic with n2g were excluded: 299 loci -> 282 loci

PCoA

- Systemic discrepancies



Identical rubias scores:

nano vs ion:	83.25%
nano vs nano:	87.22%
vc vs n2g:	98.48%

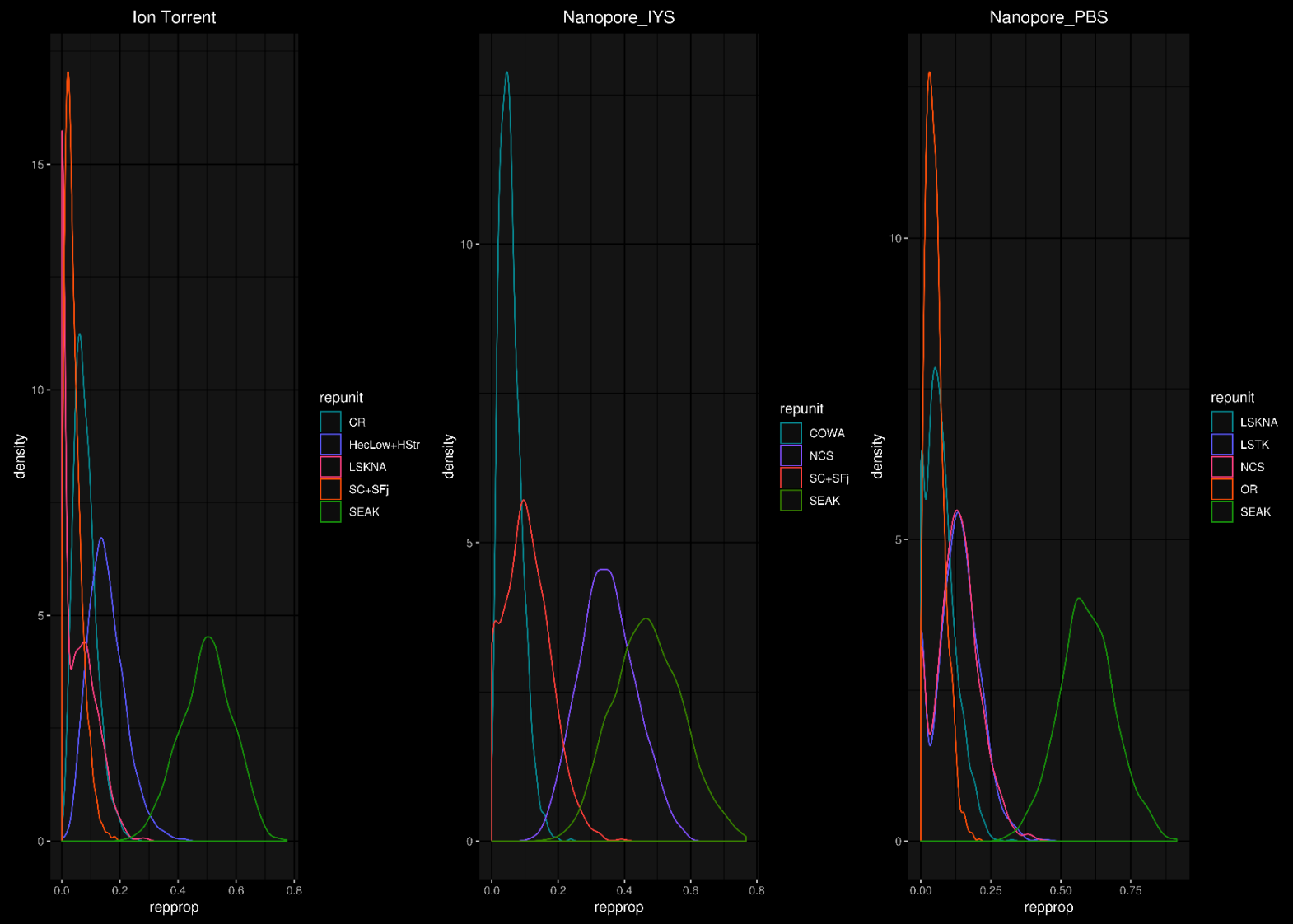
Stock assignment

Matching Reporting Unit assignment:

nano vs ion:	45.24%
nano vs nano:	40.63%
vc vs n2g:	83.67%

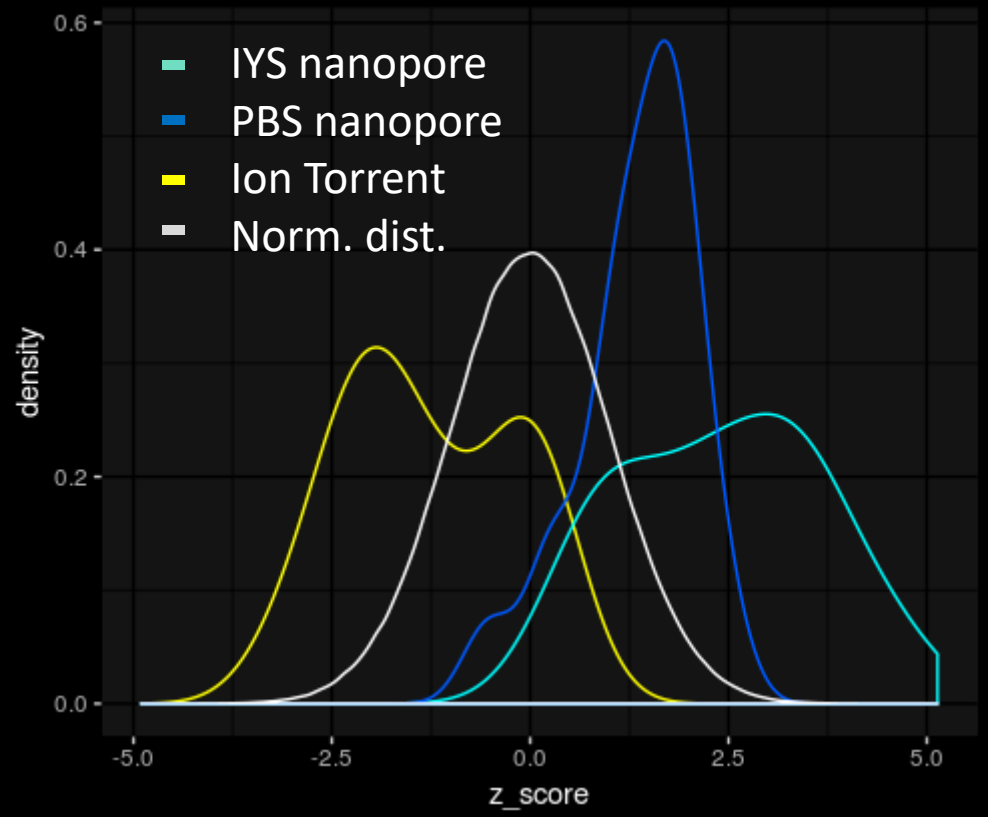
Matching Collection assignment:

nano vs ion:	24.39%
nano vs nano:	21.43%
vc vs n2g:	59.18%

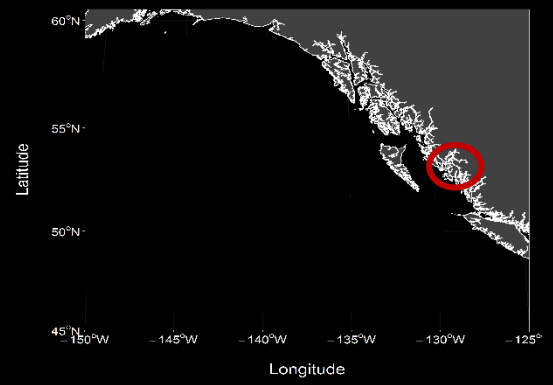


Reporting unit composition

Poor database representation!

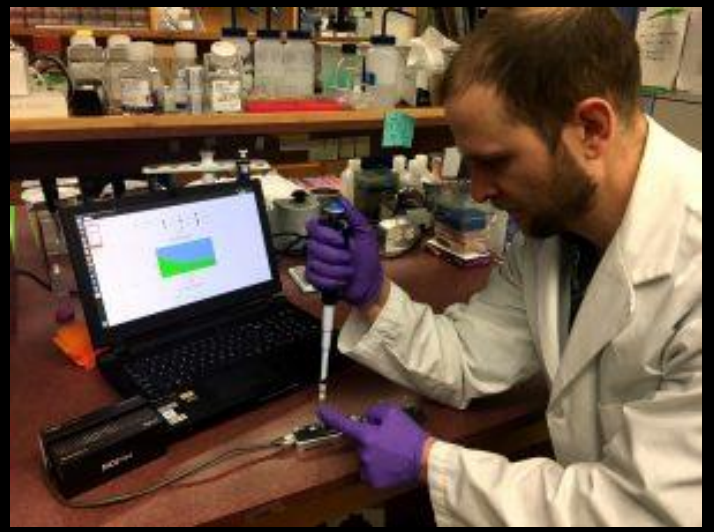


- Poor database representation + nanopore bias
  - Poor assignment scores
- New baseline/panel -> Kynoch and Mussel Inlets and Douglas Channel



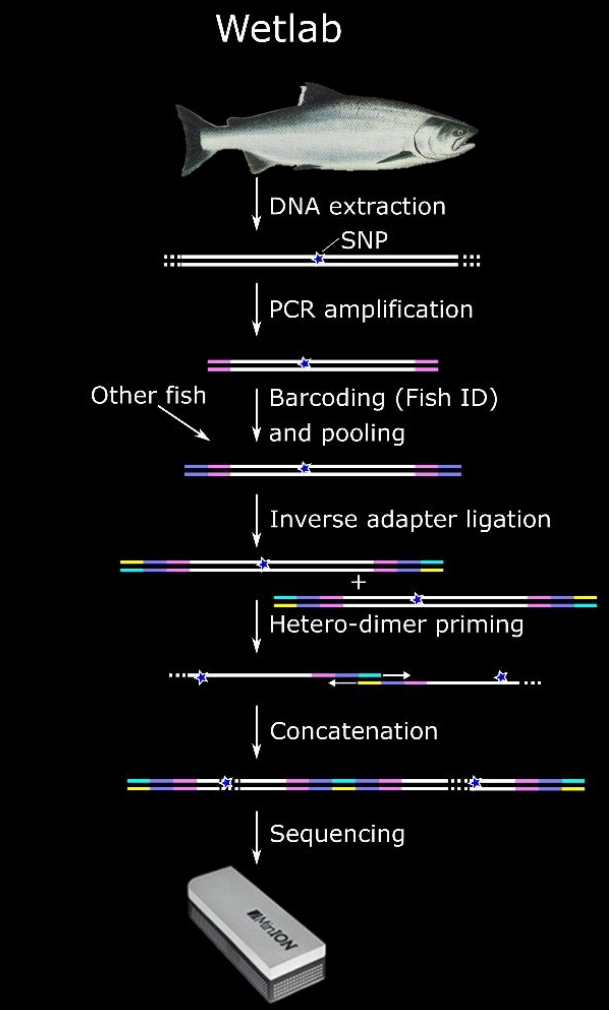
IYS nanopore GSI summary

- At sea GSI of 52/75, concordant with IonTorrent GSI
  - 83 % SNP
  - 45% Reporting unit (baseline representation)
- Turnaround
  - Wetlab 14h
  - Sequencing 12h
  - Computation: 2-3d (basecalling!!)
- Throughput:
  - ~ Max 96 fish / flow cell
- Cost: Currently 5x of Ion Torrent
- Improvements needed!



Nano2geno improvements:

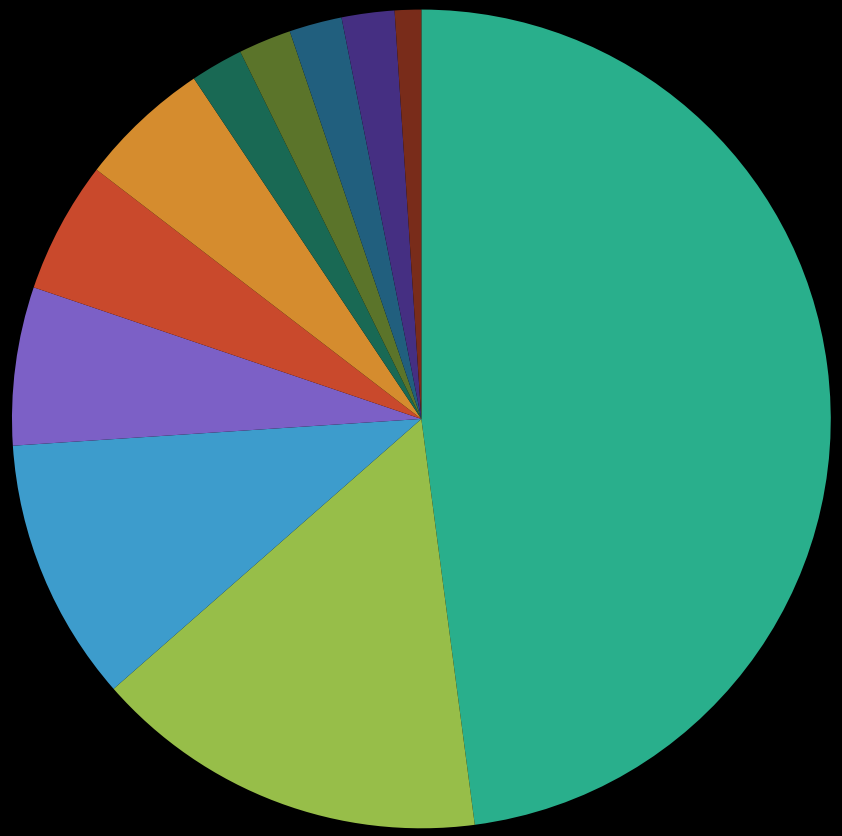
- Include barcode linker in primer
  - Faster, less none bin reads
- Improve concatenation efficiency
  - Higher throughput
- Prealiquot into 96 well plates
  - Quicker, less risk of BC cross contamination
- Use R10 flow cell
  - Lower errors -> drop coverage requirements



Nano2geno improvements

- Computational improvement
  - Use minIT -> actual real-time basecalling
- Improvement estimates:
  - < 24h from extraction to GSI
  - 5-10x throughput
  - Cut costs
  - Reuse flow cell -> variable batch size
  - Simpler protocol

GoA Coho Stock composition



SEAK	46
MusKyn	15
NCS	10
DOUG	6
CR	5
LSKNA	5
OR	2
EVI+GStr	2
WVI	2
COWA	2
HecLow+HStr	1

Thank you!

Prof. Kaganovsky crew

IYS team

Molecular Genetic Lab

Ben Sutherland , Tobi Ming, Collin Wallace, Kim Jonsen, Kelsey Flynn, Terry Beacham

Richard Beamish