

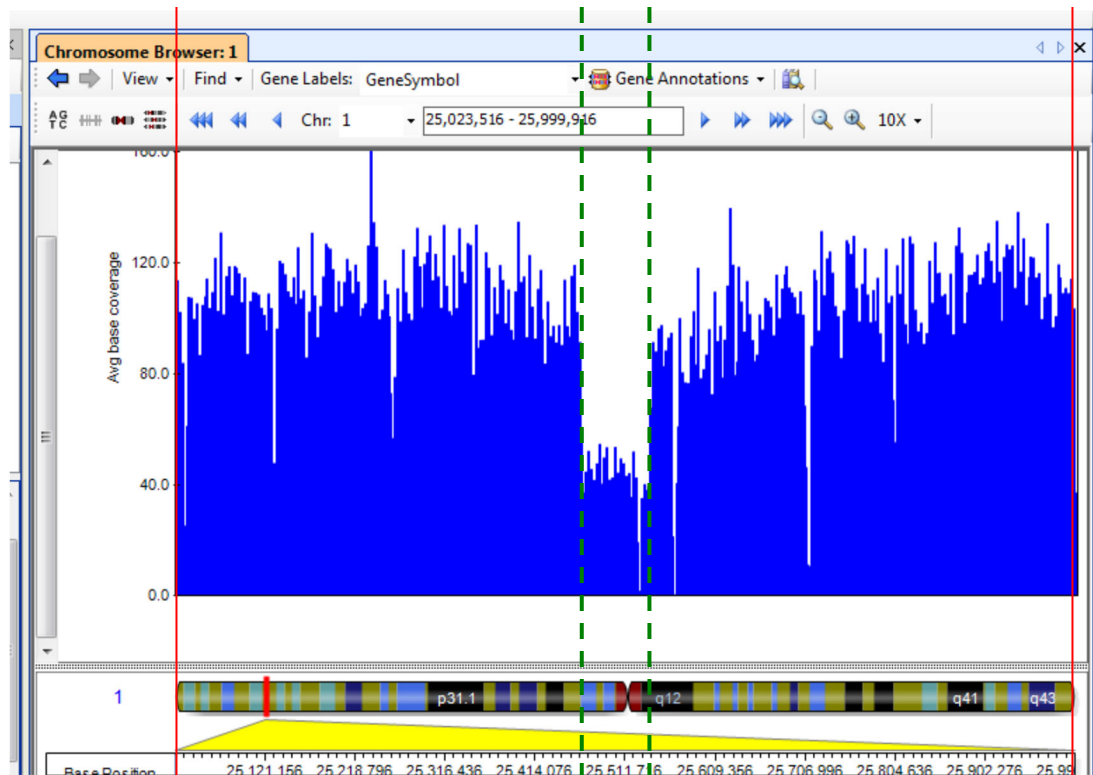
Incoming Data QC and our ability to detect commonly known genetic diseases

SOME WELL KNOWN MENDELIAN DISEASES

Albinism	1 in 17,000 per Wikipedia
Becker Muscular Dystrophy	3 - 6 of 100,000 per Wikipedia
Beta Thalassemia	Rates vary enormously by geography, but some popul
Canavan Disease	1 in 40 of Ashkenazi Jewish background
Charcot-Marie-Tooth Disease	36 per 100,000 per Wikipedia
Creutzfeldt-Jacob Disease	1 per million per Wikipedia
Cystic Fibrosis	1 per 4,000 per Wikipedia
Down Syndrome	1 in 733 per Wikipedia
Duchenne Muscular Dystrophy	1 in 4,000 per Wikipedia
Fabry Disease	1 in 3,100 to 4,600 per OMIM
Factor V Leiden	5% of Europeans are het, per Wikipedia
Familial Adenomatous Polyposis	1 in 10,000 to 15,000 per Wikipedia
Familial Alzheimer Disease	About 5% of Alzheimer cases
Fragile X syndrome	1 in 3,600 males; 1 in 4,000 - 6,000 females, per Wikip
Gaucher Disease	1 in 40,000 but much higher in Ashkenazi Jews, per Wi
Haemophilia (A&B)	1 in 5,000 - 10,000 per Wikipedia
Hemochromatosis	2 in 1,000 per Wikipedia
Huntington's	5-10 per 100,000 per Wikipedia
Hypertrophic Cardiomyopathy	0.2 - 0.5% per Wikipedia
Long QT Syndrome	1 in 5,000 per About.com
Marfan Syndrome	1 in 3,000 to 5,000 per Wikipedia
Paget Disease	1.5 - 8% per Wikipedia
Phenylketonuria	1 per 15,000 per Wikipedia
Pompe Disease	1 per 60,000 per Wikipedia
Retinitis Pigmentosa	1 per 4,756 per a PubMed article
Sickle Cell Anemia	1 in 5,000 in USA, 1 in 500 African Americans
Stargardt Disease	1 in 8,000 to 10,000 per on-line article.
Tay-Sachs Disease	2-3% in Ashkenazi Jews, Irish, Cajun; 0.3% otherwise
Thrombocytopenia	Complicated - TBD
Usher Syndrome	1 in 23,000 per Wikipedia
Von Willebrand Disease	1 in 10,000 have symptoms per Wikipedia
Xeroderma Pigmentosum	1 in 250,000 per Wikipedia

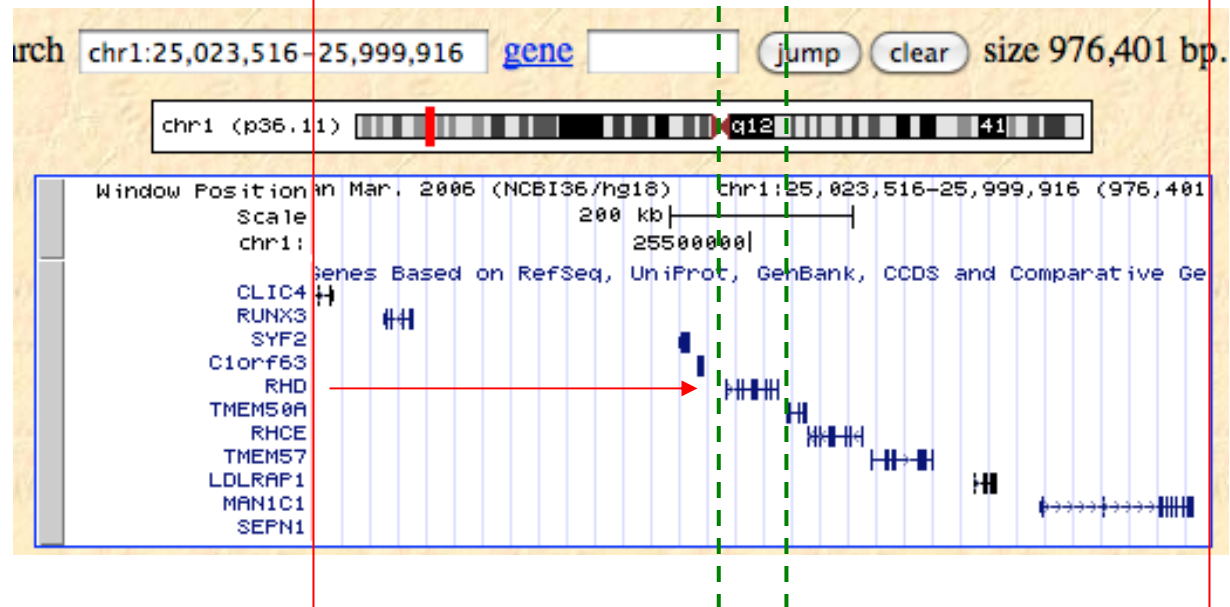
GENETIC DISEASES WHICH MAY BE CHALLENGING FOR NEXT-GEN SEQUENCING DIAGNOSIS

DISEASES	TYPE OF CHALLENGE	
Down's syndrome	CNV	CNV of all or part of Chr 21
Charcot-Marie-Tooth Disease	CNV	Duplication of a large region in Chr 17p22, including the PMP22 gene
Breast cancer	CNV	PRIM2 susceptibility gene is in a compression
Rh factor (ABO blood grouping)	CNV	Het or homozygous deletion of RHD gene
Beta Thalassemia	Pseudogenes	HBB gene has multiple pseudogenes
Gaucher Disease	Pseudogenes	GBA gene has at least one nearby pseudogene
Drug metabolism	Pseudogenes	CYP450 family has many (58?) pseudogenes (e.g. CYP3A5P1 for CYP3A5)
Fragile X syndrome	Tri-nucleotide repeat	Tri-nucleotide repeat (normally 6-55 becomes 230-4000) in FMR1
Huntington's	Tri-nucleotide repeat	Variable tri-nucleotide repeat in the HTT gene, on Chr 4 (28 - >60)
Pompe Disease	High GC content	GAA gene is about 80% GC
Marfan Syndrome	High de novo rate	Mutations of FBN-1 gene: 15-30% are de novo



Coverage Graph Shows Medically Relevant Large Het Deletion

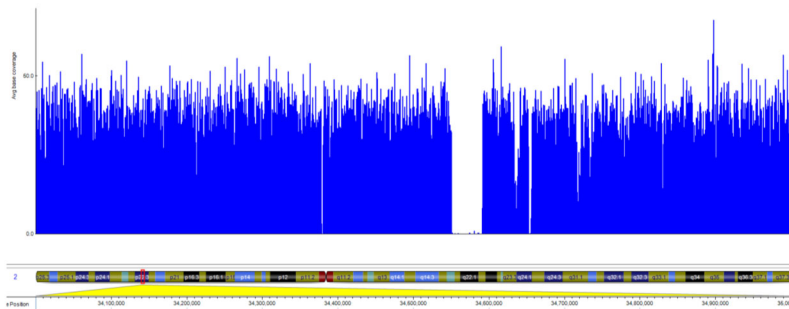
Heterozygous deletion of approx. 70kbp on Chr 1 completely eliminates one copy of the RHD gene. This deletion, heterozygous in 40% of Europeans, and homozygous in 16%, is responsible for Rh-negative ABO blood typing when homozygous.



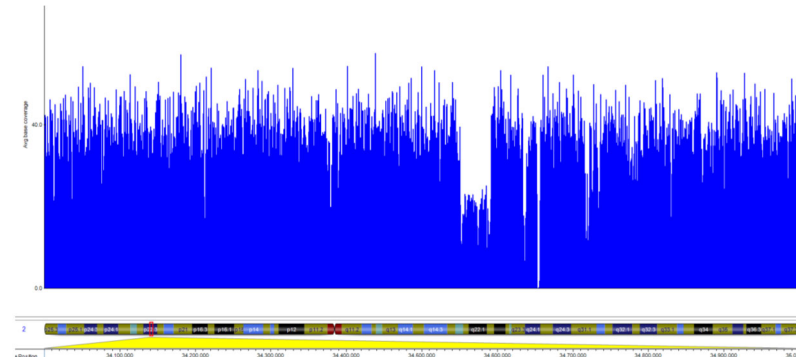
Large Het or Hom Deletions Can Be Seen From Coverage

Example: 40kbp Het / Hom Del in Chr 2

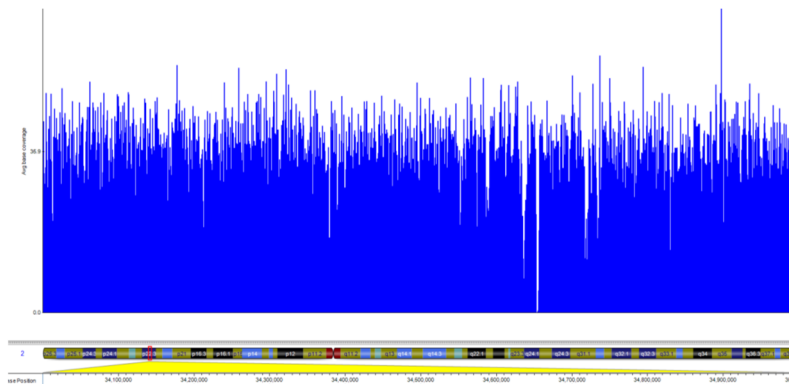
West family genome data, Chr2 34-35M base range



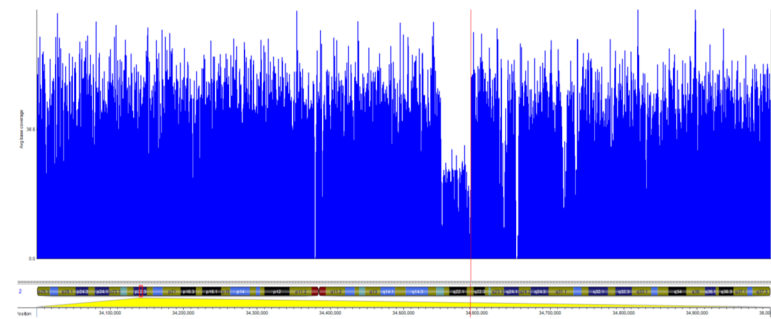
Father



Daughter



Mother



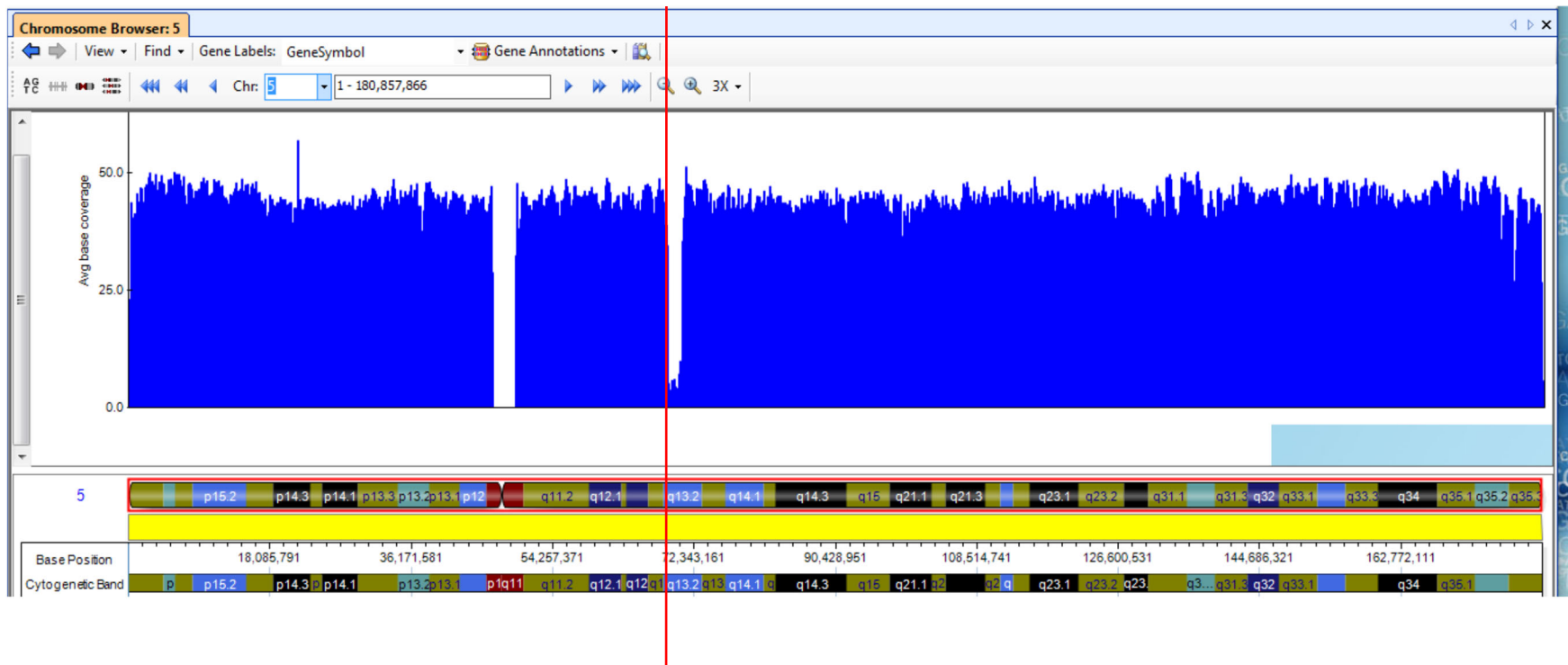
Son

(This particular example is inside a compression, so het SNP's are observed inside a het deletion)

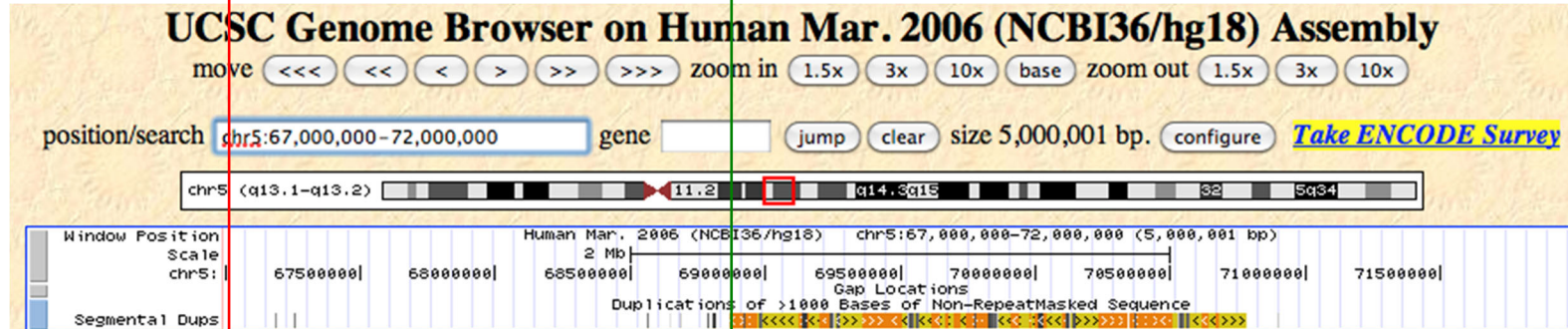
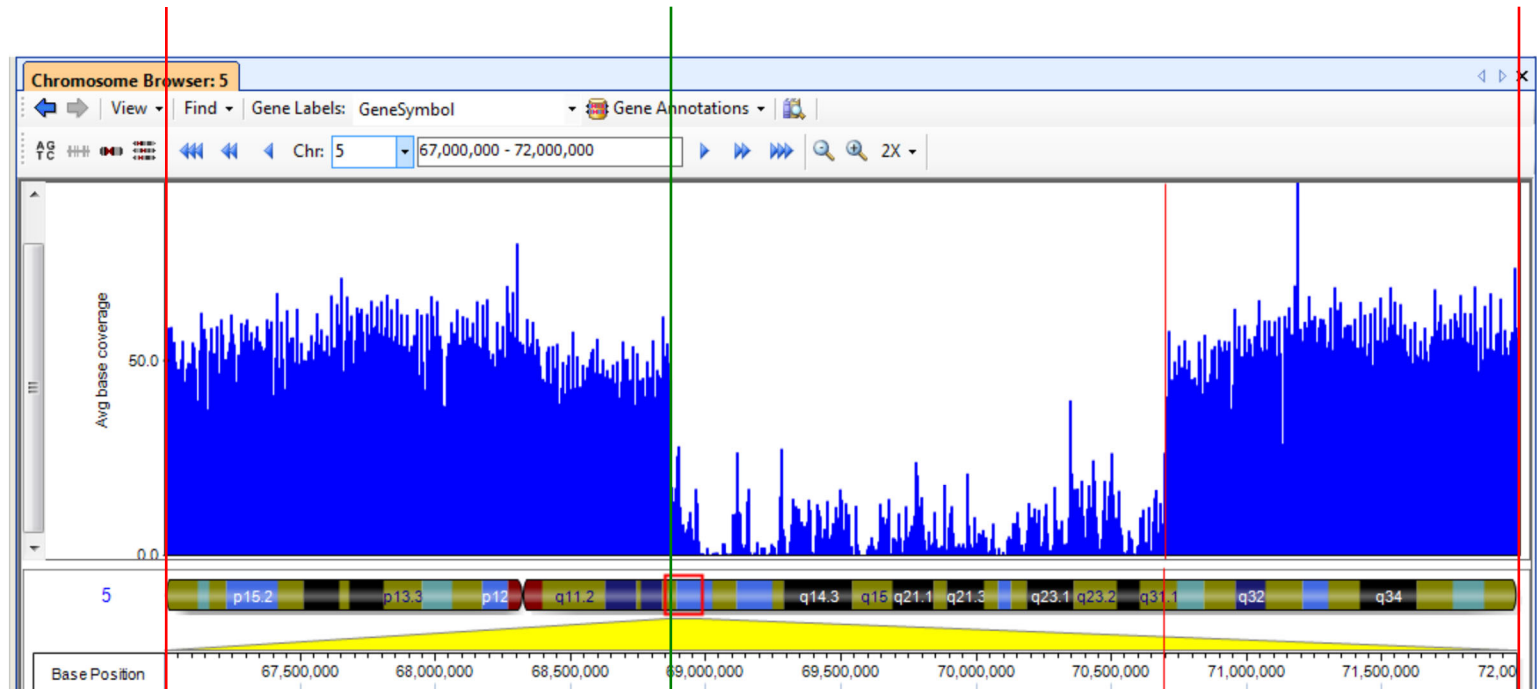


Coverage Distribution Over All of Chromosome 5

Drop-Outs Due to Centromere and (Red line) Segmental Duplication
Zoomed-In on Next Slide



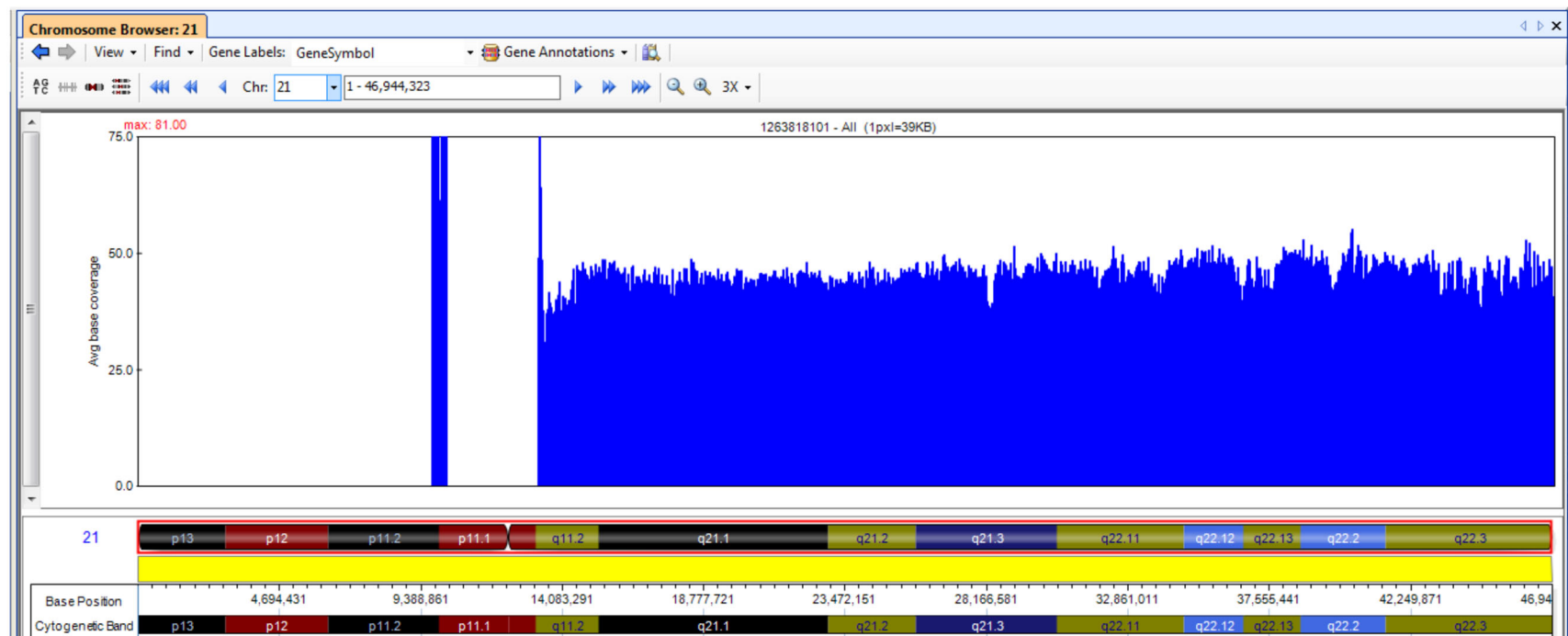
1.85 Mbp Coverage Drop-Out on Chr 5 Due To Segmental Duplication



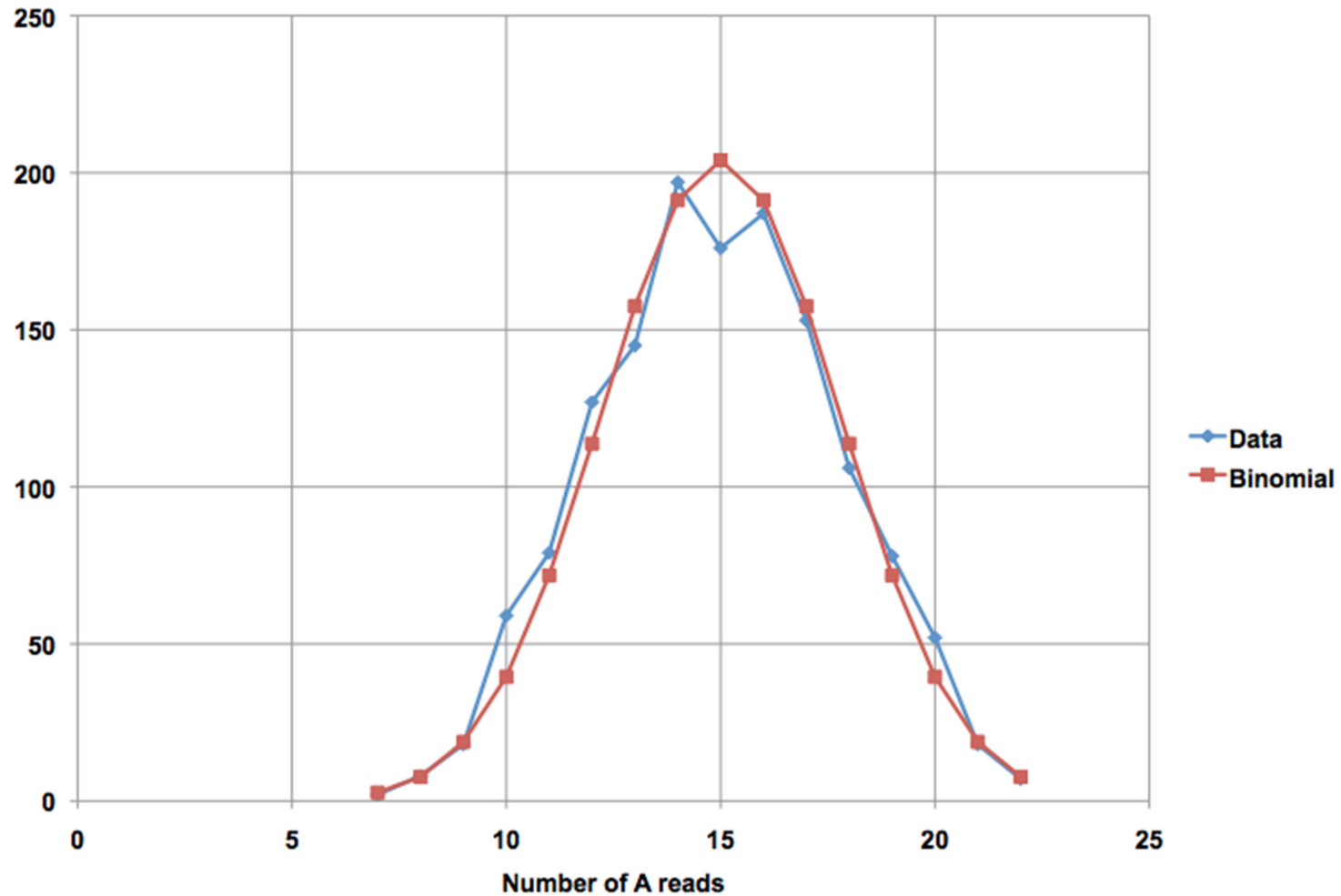
Coverage Over All of Chromosome 21 – Potential to Detect Down's Syndrome

Trisomy over part of the chromosome should be clear from this graph.

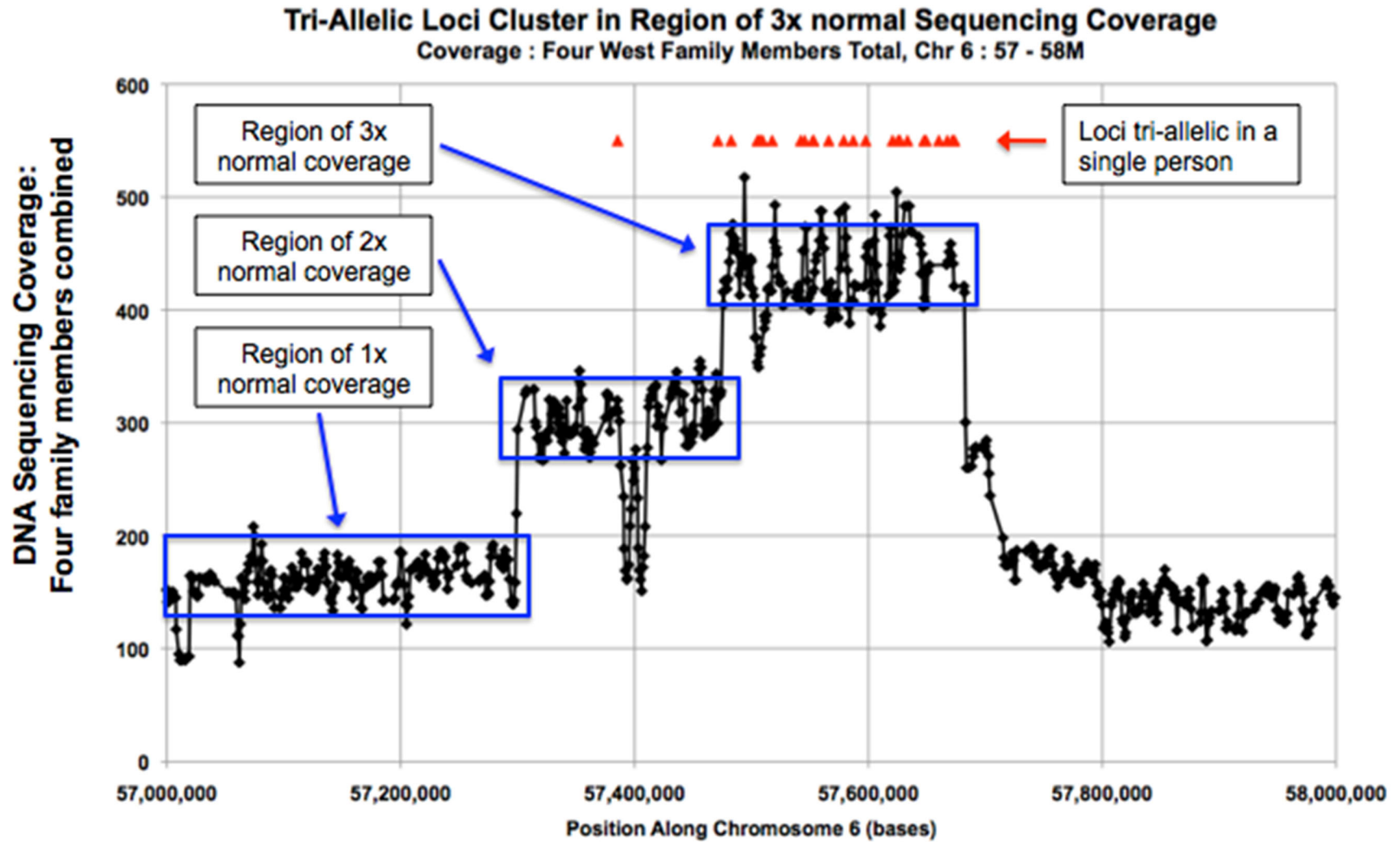
Trisomy over all of the chromosome should be clear from chromosome coverage ratios

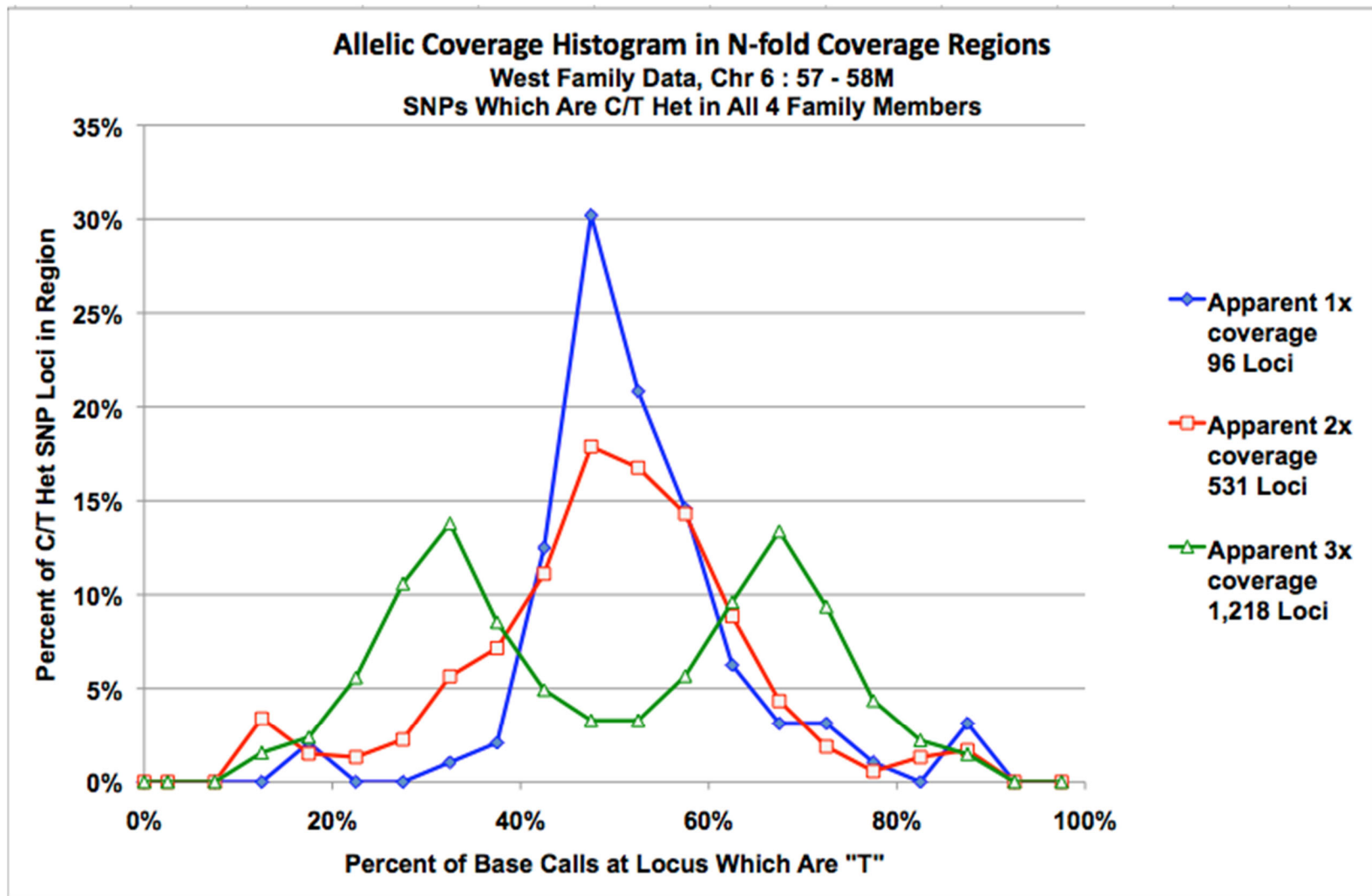


AG or GA Het SNP's : Histogram of # A reads - Data vs Binomial
At all SNP loci selected, the total number of reads was 30



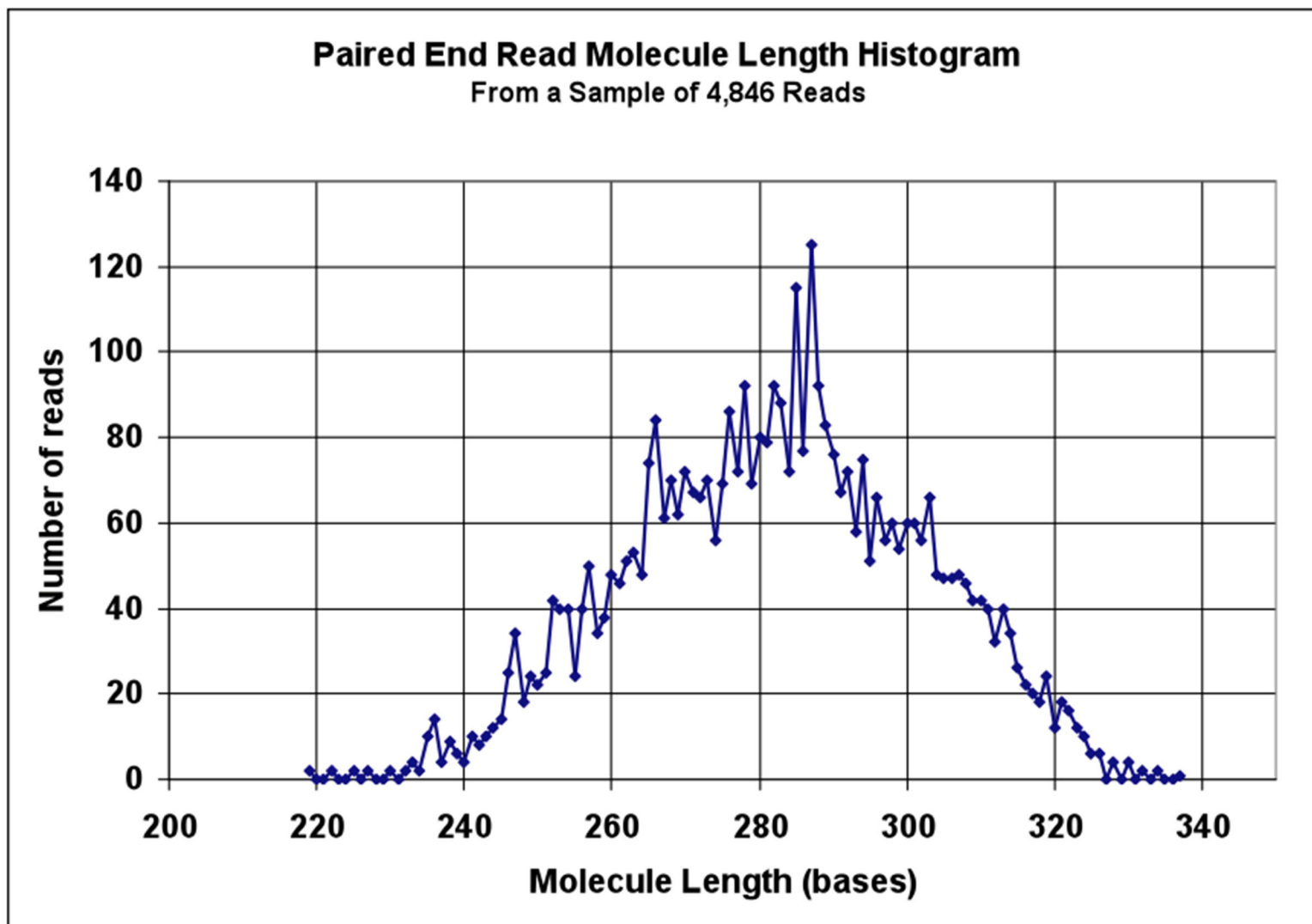
AG or GA het SNP loci from Chromosome 1 of Anne's genome sequence, at which exactly 30 reads were used to make the call. This shows that the allelic balance matches the expected binomial probability distribution quite closely.





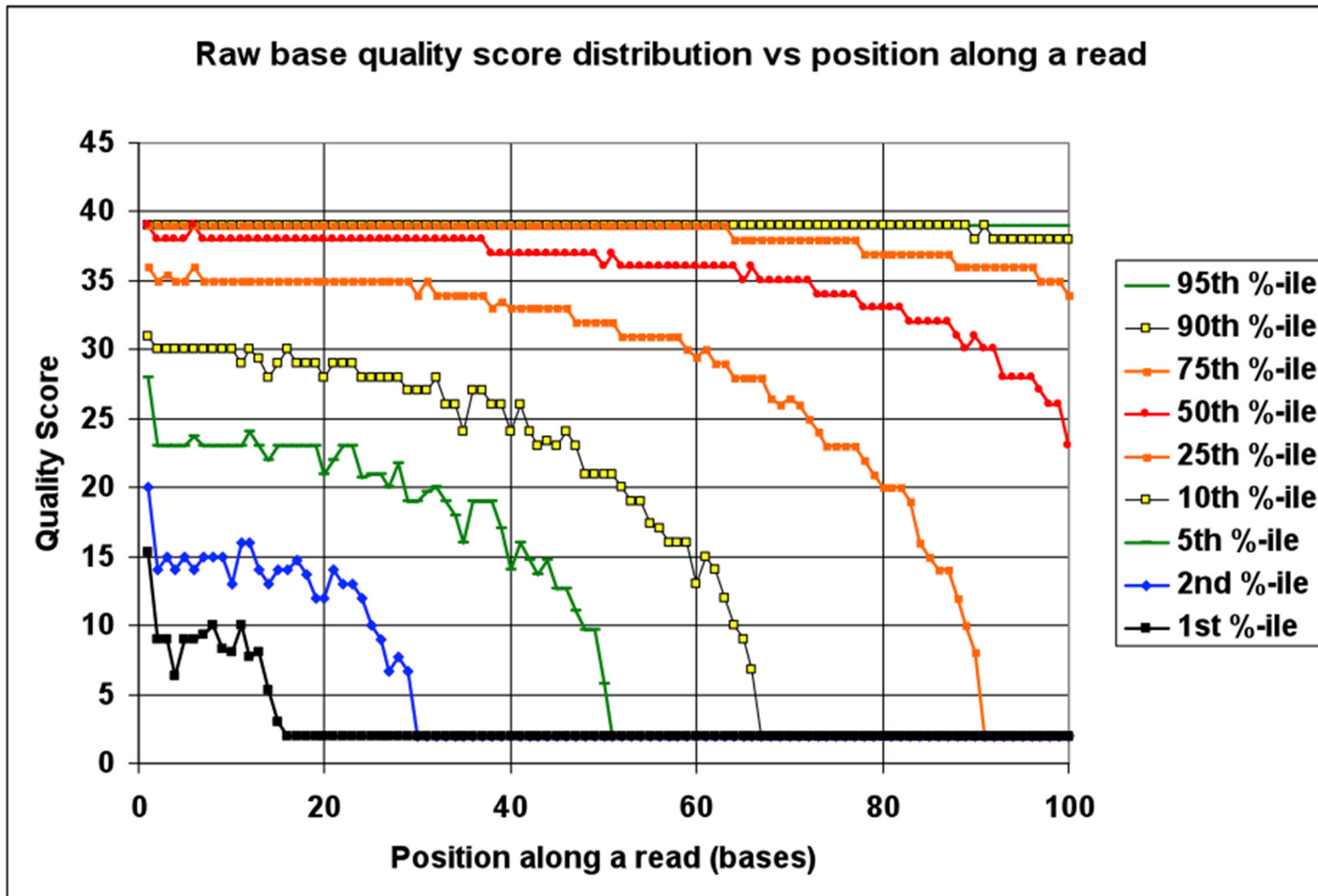
The region which appears to have 3x nominal overall coverage also shows an allelic distribution which clearly confirms the 3-fold repeat structure.

Paired-End Read Length Distribution
A Simple Sample Prep QC Metric



Distribution of Raw-Read Individual Base Q-Scores Along A Read

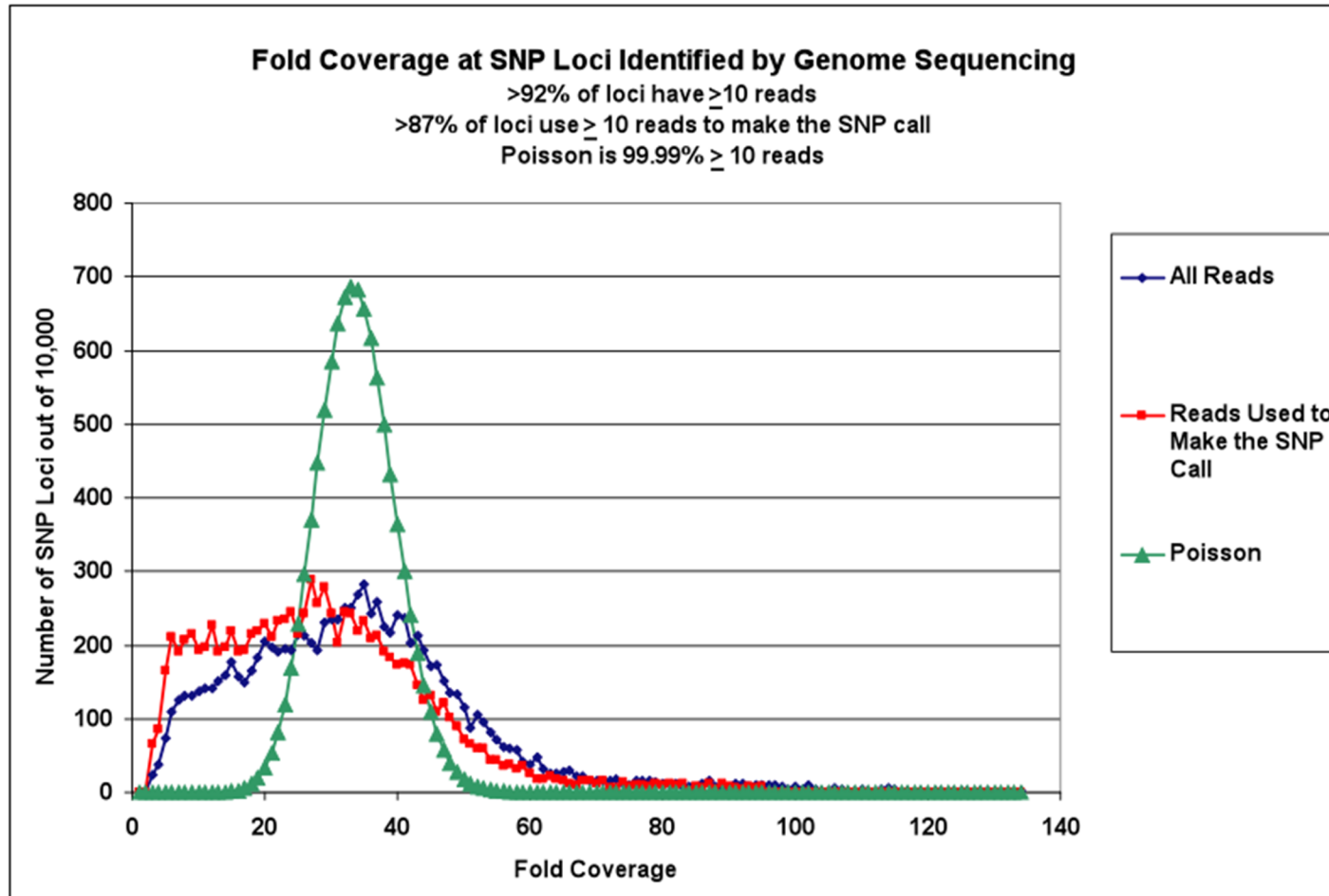
A Simple Sequencing QC Metric



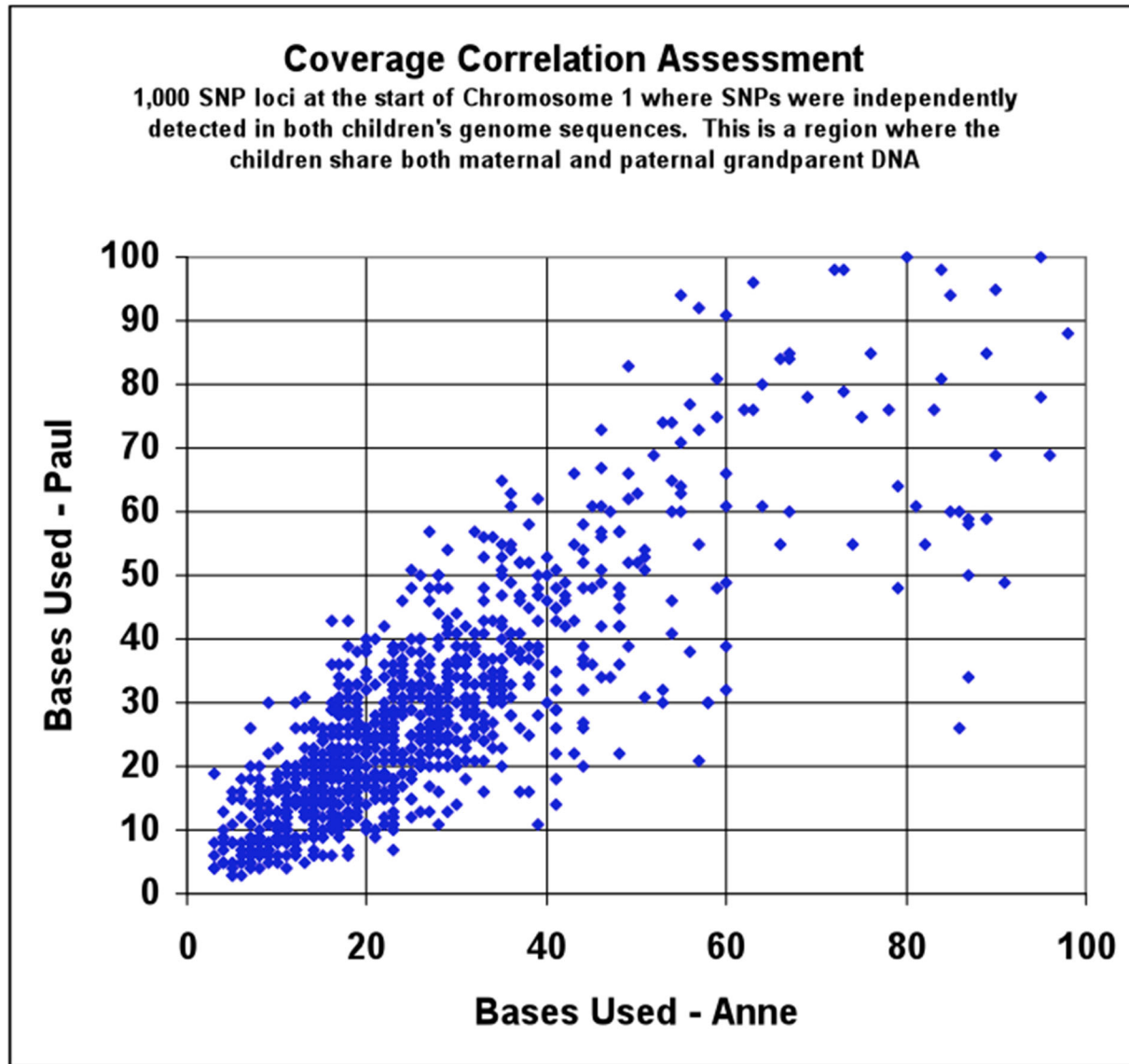
Coverage Histogram vs Poisson

A Simple Sequencing QC Metric

In this example, 13% of SNP Genotypes Were Based on < 10-fold Coverage

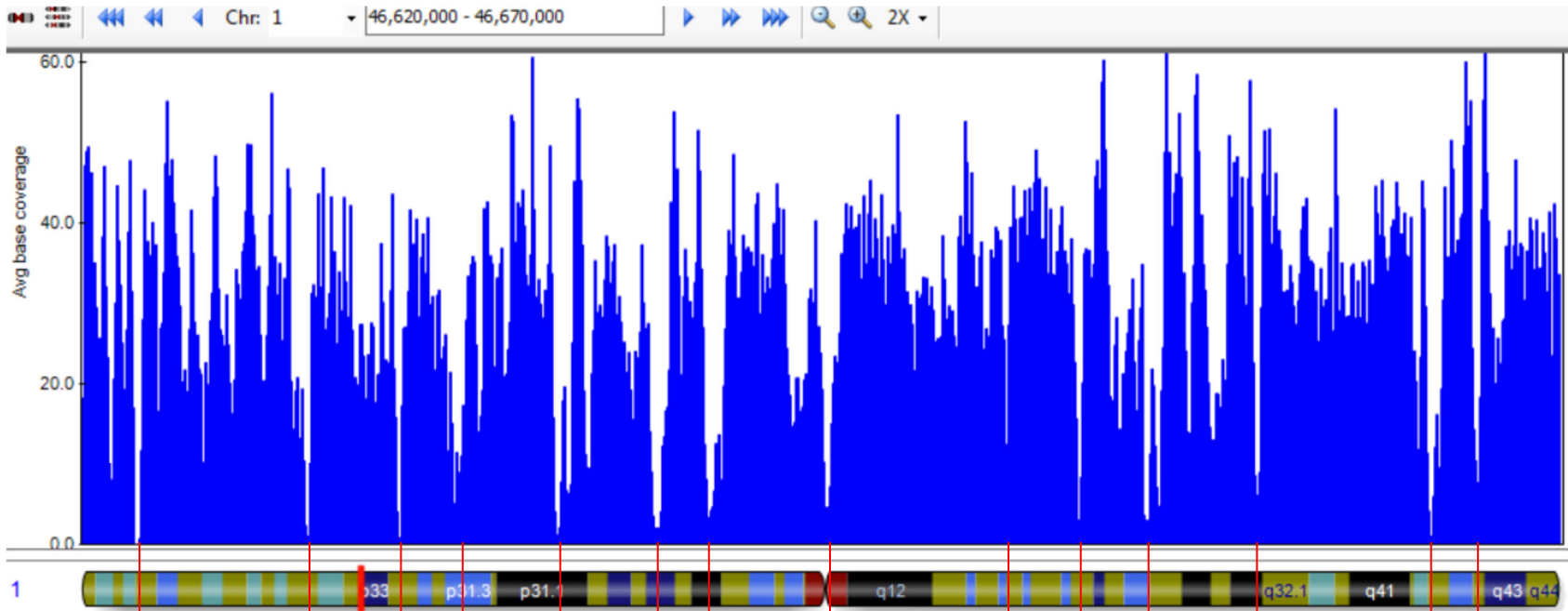


Coverage Is Significantly Correlated Between Individuals

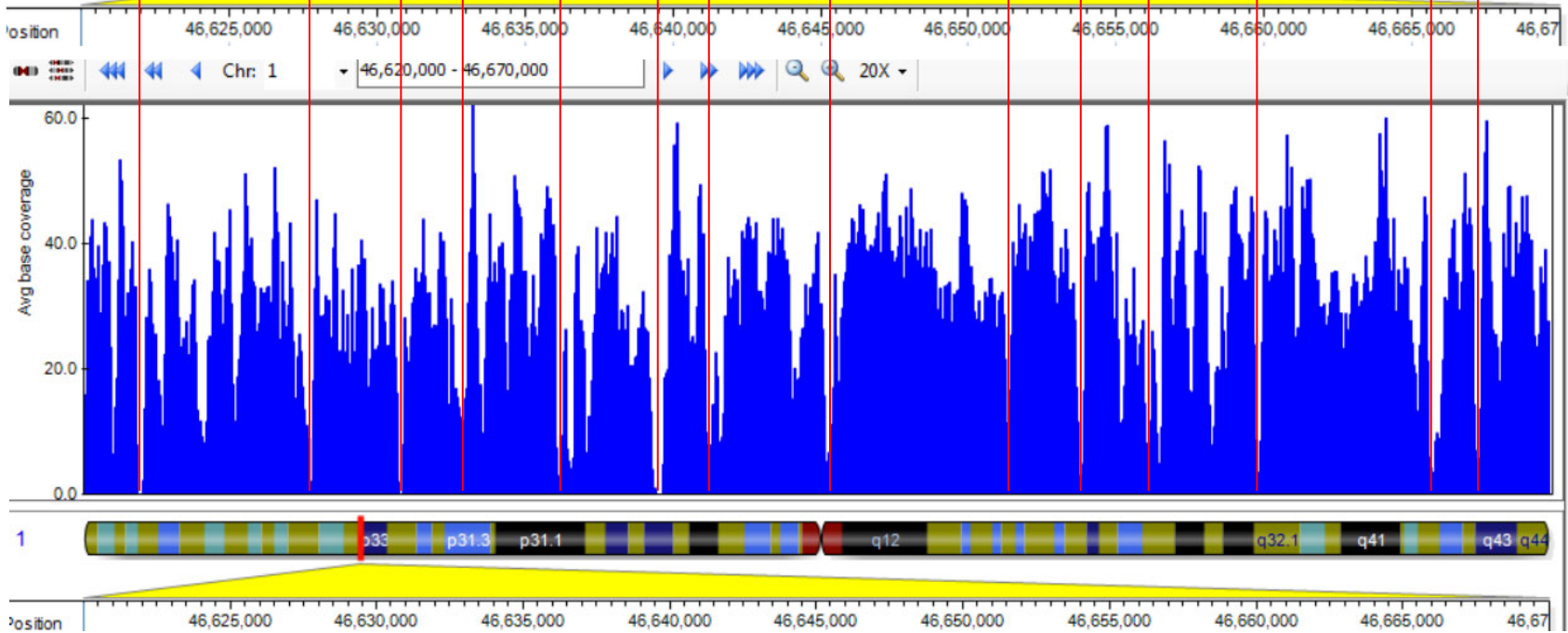


Comparison of coverage dips / gaps in the same region of
two genetically unrelated people

John

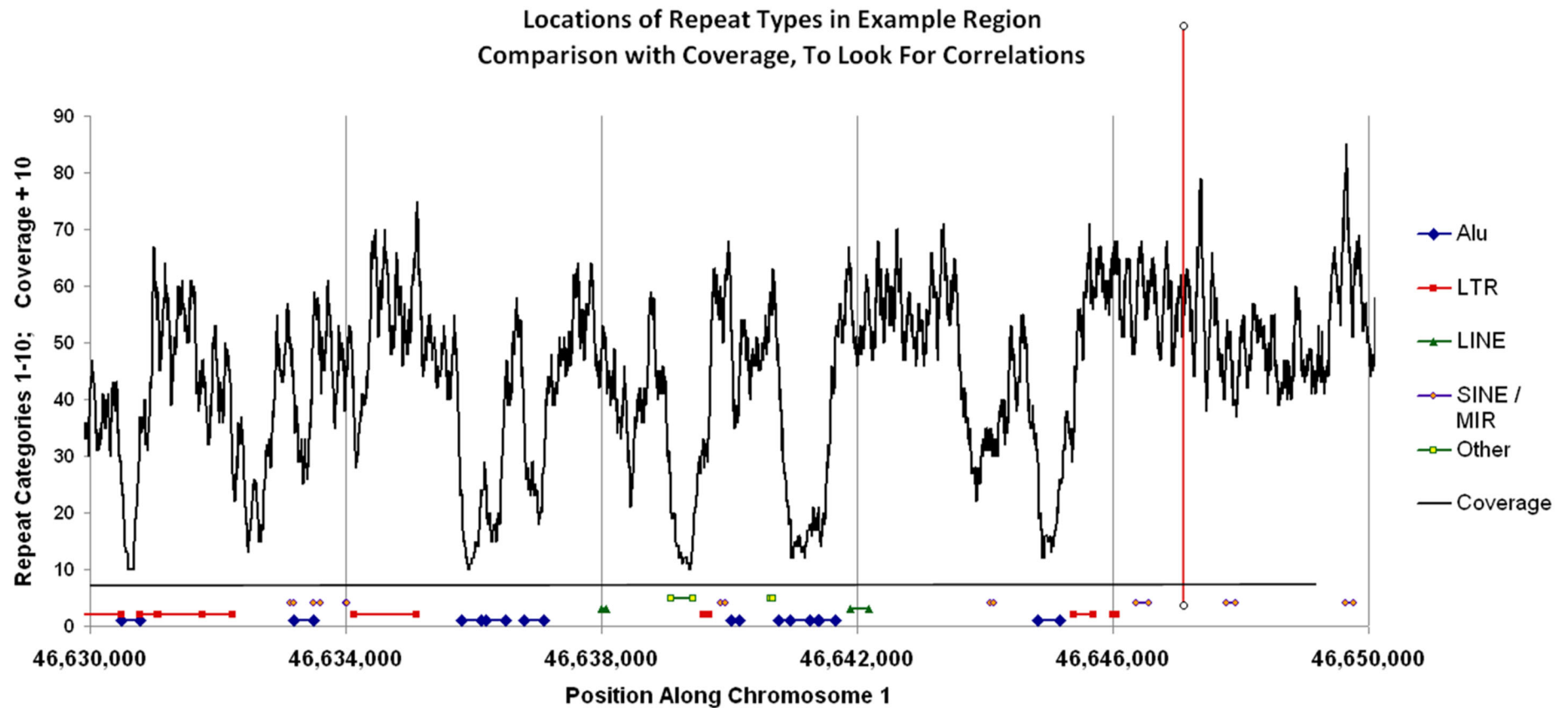


Judy



Coverage Variation Correlated With Alu Repeats

Illumina Alignment From Early 2010

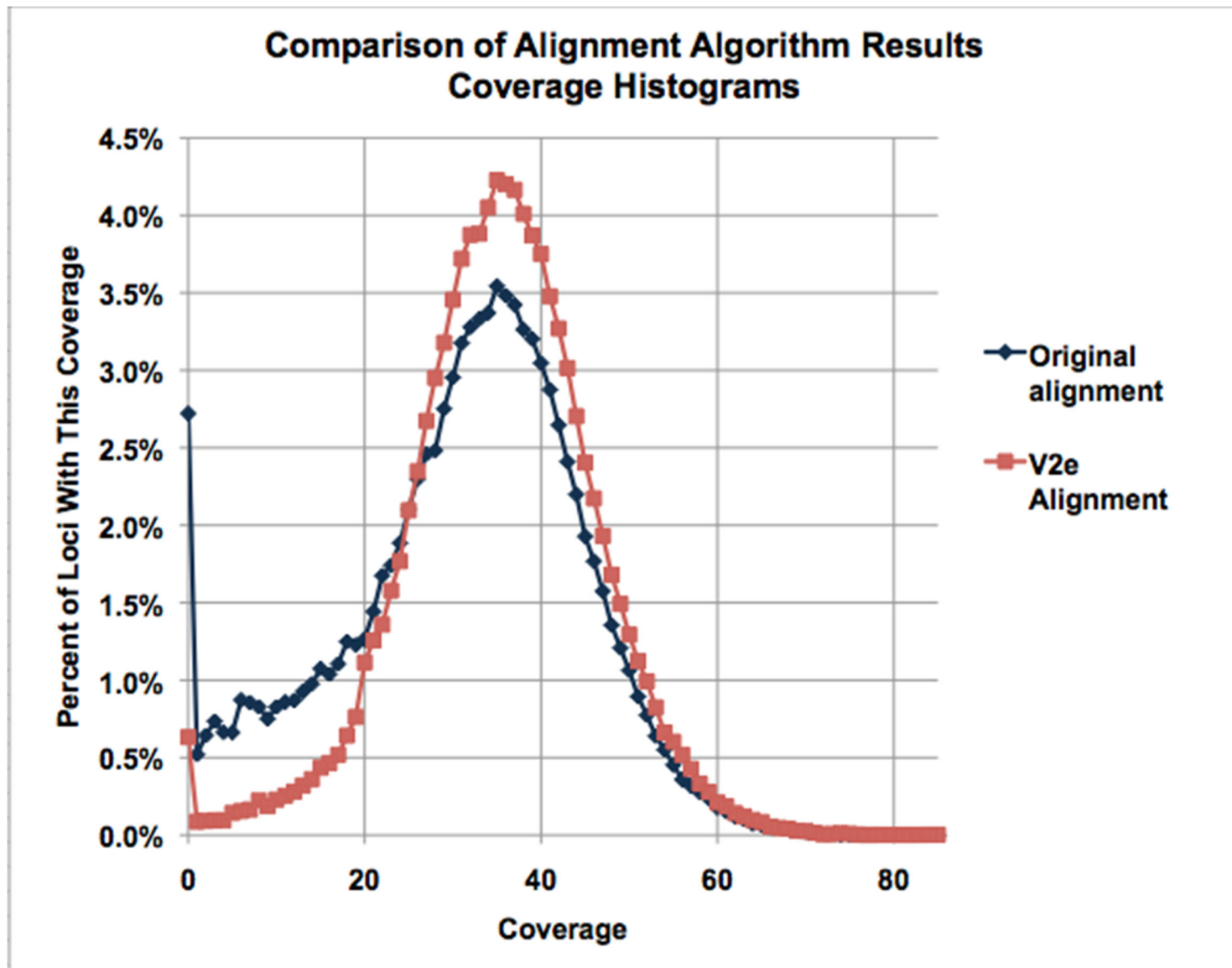


Initial results with the new Illumina V2e Alignment Software

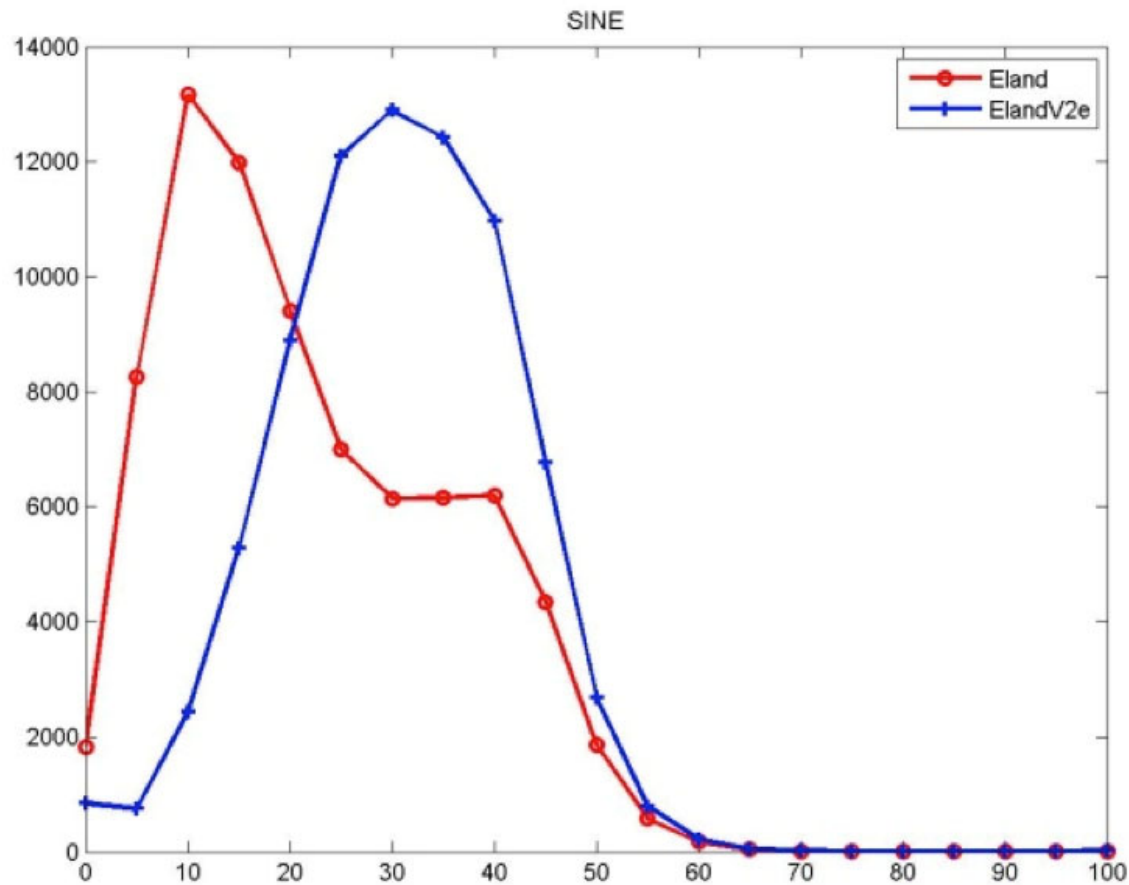
Based on a West Family Genome Aligned Using V2e

John & Anne West

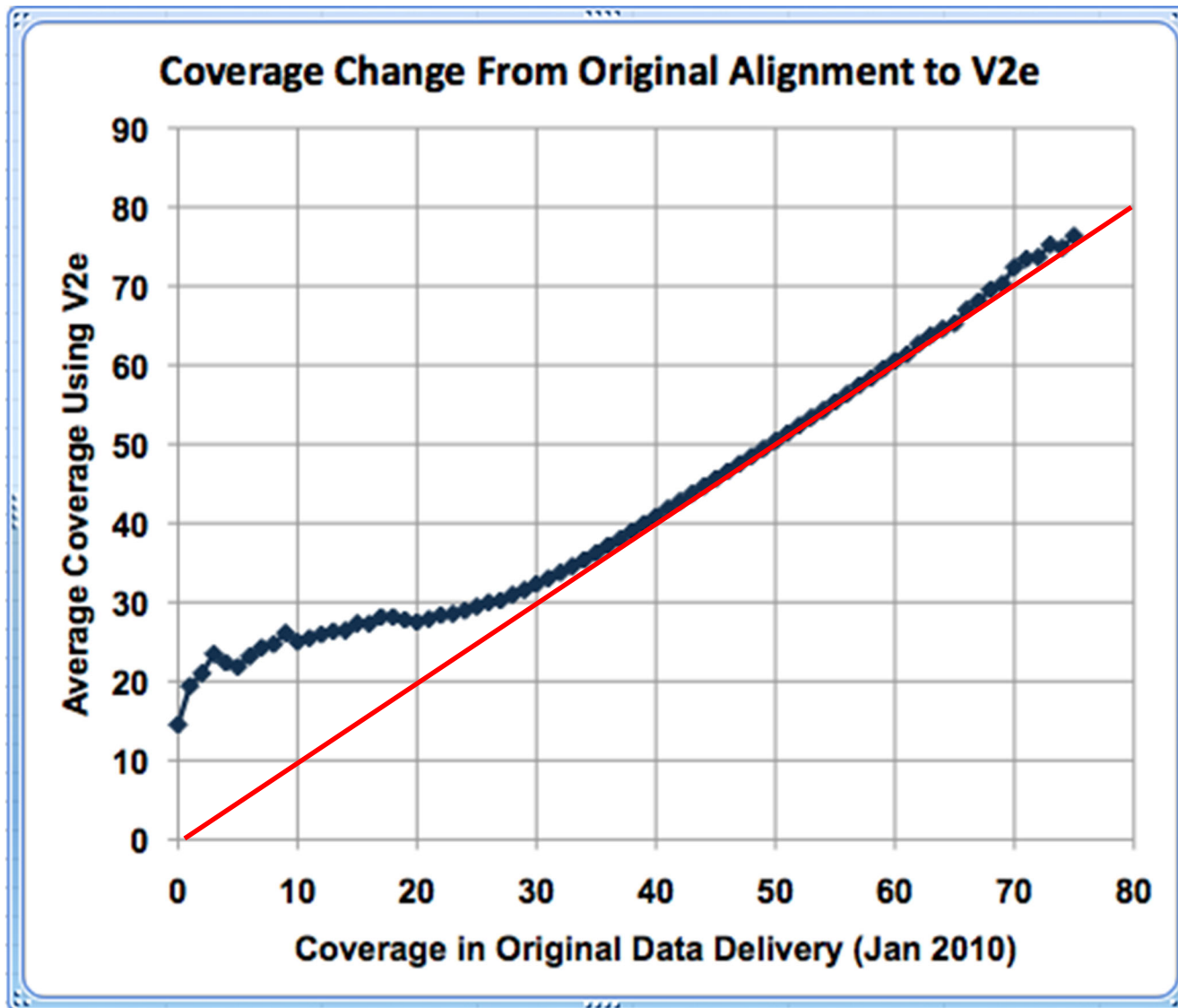
Oct 12, 2010



Comparison of coverage histograms in a “typical” region of 250,000 bases starting at position 57,000,000 on Chromosome 6.

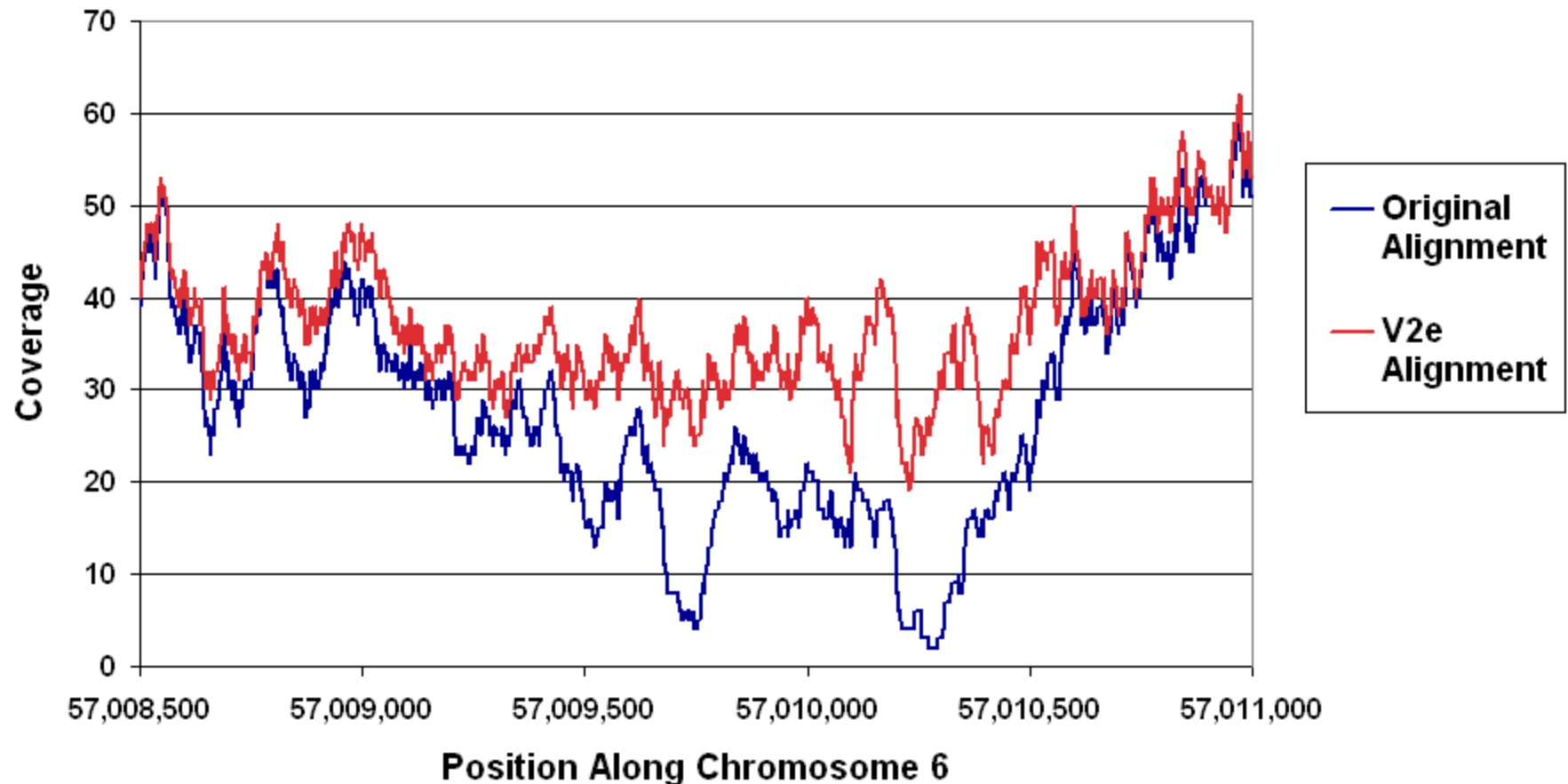


Coverage becomes much closer to Poisson with the Illumina Eland v2e algorithm, particularly in short repeats (SINE elements) e.g. Alu repeats.



Most of the coverage increase is in regions where the original alignment yielded low coverage. Loci which had 10-fold coverage in the original alignment, for example, averaged about 25-fold coverage in the V2e alignment.

DNA Sequencing Coverage Comparison



Example comparing DNA sequencing coverage from the original alignment of one of our genomes with the same region after use of the new V2e alignment software.