www.ScienceTranslationalMedicine.org 12 January 2011 Vol 3 Issue 65 65ra4

Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing

Callum J. Bell,[1]* Darrell L. Dinwiddie,[1,2]* Neil A. Miller,[1,2] Shannon L. Hateley,[1] Elena E. Ganusova,[1] Joann Mudge,[1] Ray J. Langley,[1] Lu Zhang,[3] Clarence C. Lee,[4] Faye D. Schilkey,[1] Vrunda Sheth,[4] Jimmy E. Woodward,[1] Heather E. Peckham,[4] Gary P. Schroth,[3] Ryan W. Kim,[1] Stephen F. Kingsmore[1,2†]

[1]National Center for Genome Resources, Santa Fe, NM 87505, USA. [2]Children's Mercy Hospital, Kansas City, MO 64108, USA. [3]Illumina Inc., Hayward, CA 94545, USA. [4]Life Technologies, Beverley, MA 01915, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: sfk@ncgr.org

Presented by: Baikang Pei
Journal Club, Gerstein Lab, 2011-01-25

Tay-Sachs Disease (TSD)

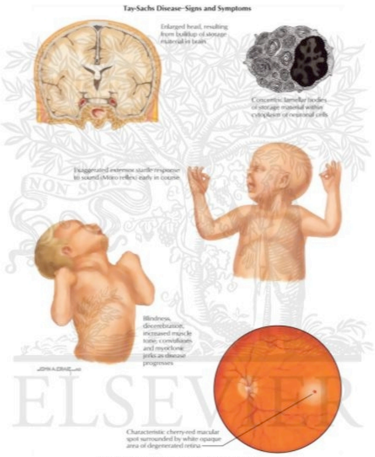
TSD: an autosomal recessive neurodegenerative disorder

Onset of symptoms in infancy and death by 2 to 5 years of age

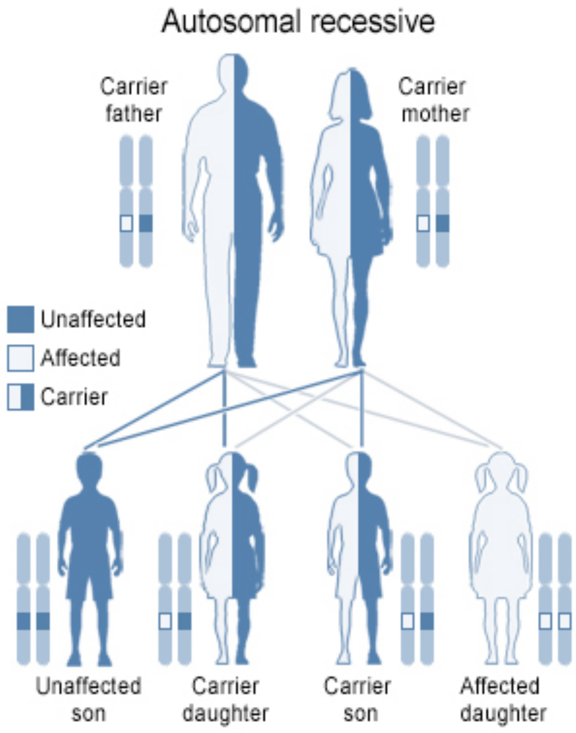
Premature death of nerve cells of the brain due to gangliosides accumulation

TSD is incurable, but treatments are available

Affected couples may decide not to have children or to conceive a child using IVF treatment


Tay-Sachs Disease–Signs and Symptoms
Enlarged head, resulting from buildup of storage material in brain
Concentric lamellar bodies of storage material within cytoplasm of neuronal cells
Exaggerated extensor startle response to sound (Moro reflex) early in course
Blindness, decerebration, increased muscle tone, convulsions and myoclonic jerks as disease progresses
JOHN A.CRAIG
ELSEVIER
Characteristic cherry-red macular spot surrounded by white opaque area of degenerated retina

© ELSEVIER, INC. – NETTERIMAGES.COM



U.S. National Library of Medicine

Preconception Screening

Of 7028 disorders with suspected Mendelian inheritance, 1139 are recessive and have an established molecular basis.

They account for ~20% of infant mortality and ~10% of pediatric hospitalizations.

To date, preconception carrier testing has been recommended in the USA only for five diseases.

Some major obstacles

Rear disease, high cost, absence of accurate, sensitive and scalable technologies

Target Capture and NGS

Target capture and NGS are considered as a potential paradigm for carrier testing for their cost-effectiveness and broad coverage of mutations.

Target capture: to targeted and amplified particular sequences in DNA samples that were known to be associated with the recessive disease genes

Challenges

More stringent sensitivity and specificity are required for routine use in clinical practice than usual genome research

Design

Disease Inclusion

448 diseases were chosen that would almost certainly change family planning by prospective parents or affect antenatal, perinatal, or neonatal care

Genome Coverage

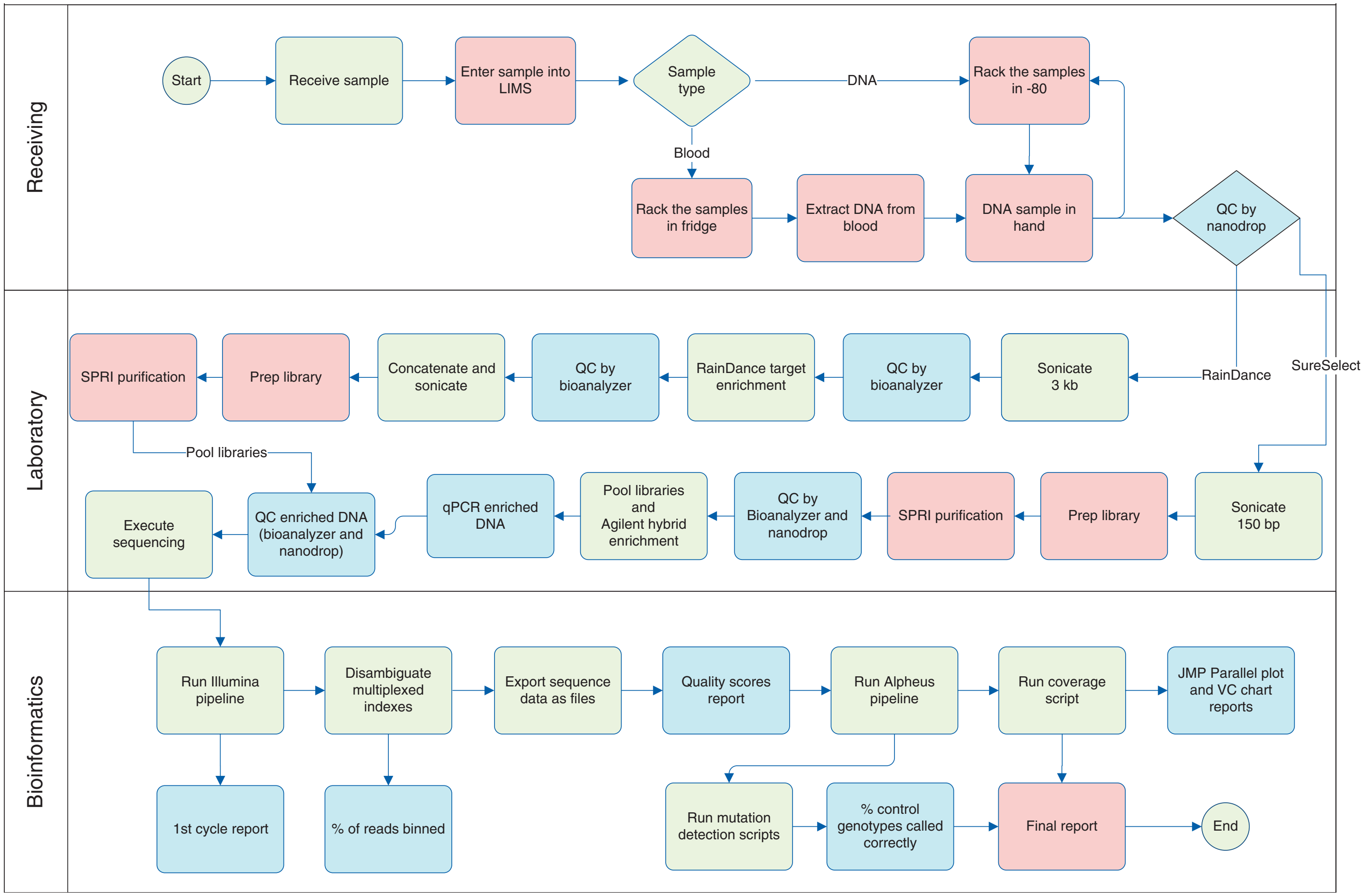
close to 2M nucleotides corresponding to 7717 segments of 437 disease genes. Targeted were exons, introns, splice junctions, regulatory regions and UTR

Samples: 104 unrelated individuals, 76 were known to be carrier or affected

Target capture: Agilent SureSelect hybrid capture, RainDance microdroplet PCR

NGS: Illumina GAIIx, SOLiD, Illumina HiSeq

Workflow



DNA sample collection

↓

Target enrichment
(RainDance, SureSelct)

↓

NGS
(GA, SOLiD, HiSeq)

↓

Analysis

Statistics

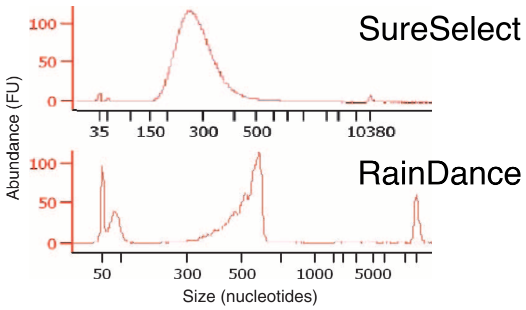
Table 1. Sequencing, alignment, and coverage statistics for target enrichment and sequencing platforms.

Sample set	Enrichment method	Sequencing method	Multi-plexing	Read length (nt)	Quality score*	Total reads ± %CV*[†]	% uniquely aligning reads*	Total nucleotides*	Aligning depth*	% nt on target ± %CV*	Fold enrichment*	% 0× coverage*	% ≥20× coverage*	Coverage ± %CV*	Pearson's coefficient[‡]
1 (*n* = 12)	SureSelect	GAIIx	12	50	30	9,952,972.5 ± 21	94	497,648,625	225	13.7 ± 3	214	4.83	61	27 ± 21	0.28
2 (*n* = 12)	SureSelect	GAIIx	12	50	30	10,127,721 ± 16	95	506,386,025	234	23.0 ± 2	358	3.66	80	50 ± 16	0.19
1 + 2 (*n* = 24)	RainDance	GAIIx	12	50	36	9,412,698 ± 30	97	470,634,900	196	29.6 ± 5	462	5.46	86	52.5 ± 33	0.23
1 + 2 (*n* = 12)	RainDance	GAIIx	12	50	31	12,807,392 ± 17	96	640,369,600	277	22.2 ± 7	346	4.62	88	56 ± 12	0.27
3 (*n* = 6)	SureSelect	GAIIx	6	50	30	19,711,735 ± 34	95	985,586,750	463	17.4 ± 3	273	1.80	86	76 ± 30	0.14
3 (*n* = 6)	SureSelect	SOLiD 3	6	50	24	16,506,076 ± 5	82	825,303,800	310	19.5 ± 7	304	6.08	79	58 ± 7	0.24
4 (*n* = 72)	SureSelect 2	HiSeq	8	149[§]	42[§]	9,273,596 ± 24	98	1,390,464,487	495	31.7 ± 4	494	2.33	92	152 ± 26	0.02
5 (*n* = 8)	SureSelect	HiSeq	8	149[§]	41[§]	9,861,765 ± 35	97	1,493,946,141	517	28.4 ± 4	442	2.25	93	139 ± 40	0.06

*Median value. [†]Coefficient of variation (%). [‡]Pearson's median skewness coefficient [3(mean − median)/SD]. [§]After assembly of forward and reverse 130-bp paired reads.

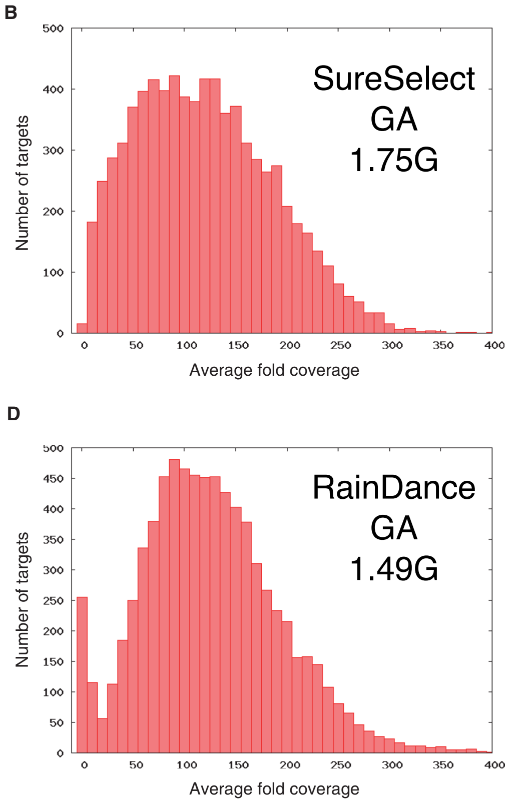
Enrichment Techniques

Distribution of size of sequencing libraries after target enrichment

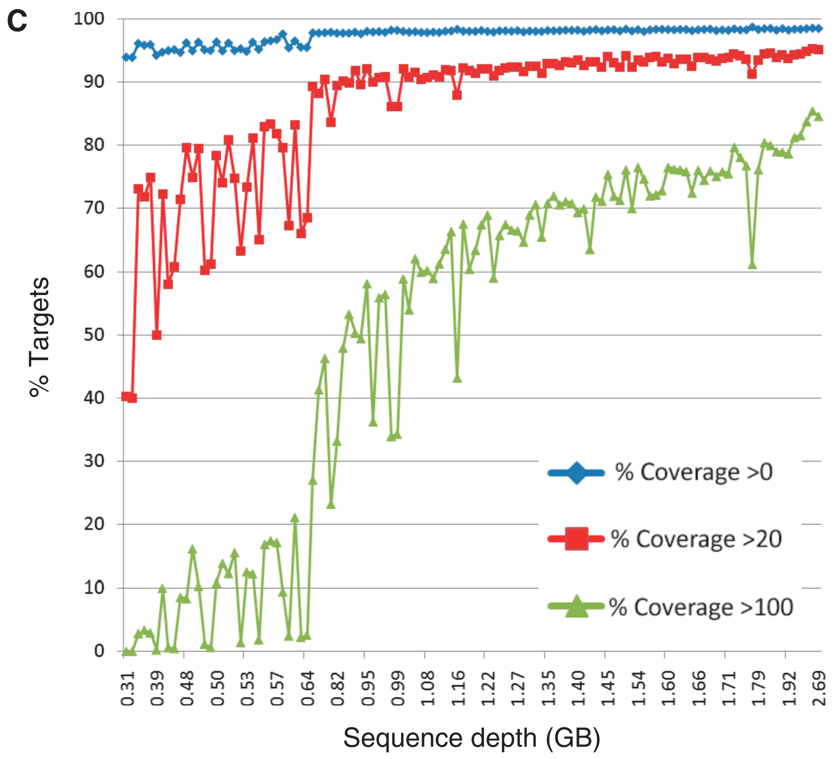


Use SureSelect in the subsequent studies

Distribution of target coverage



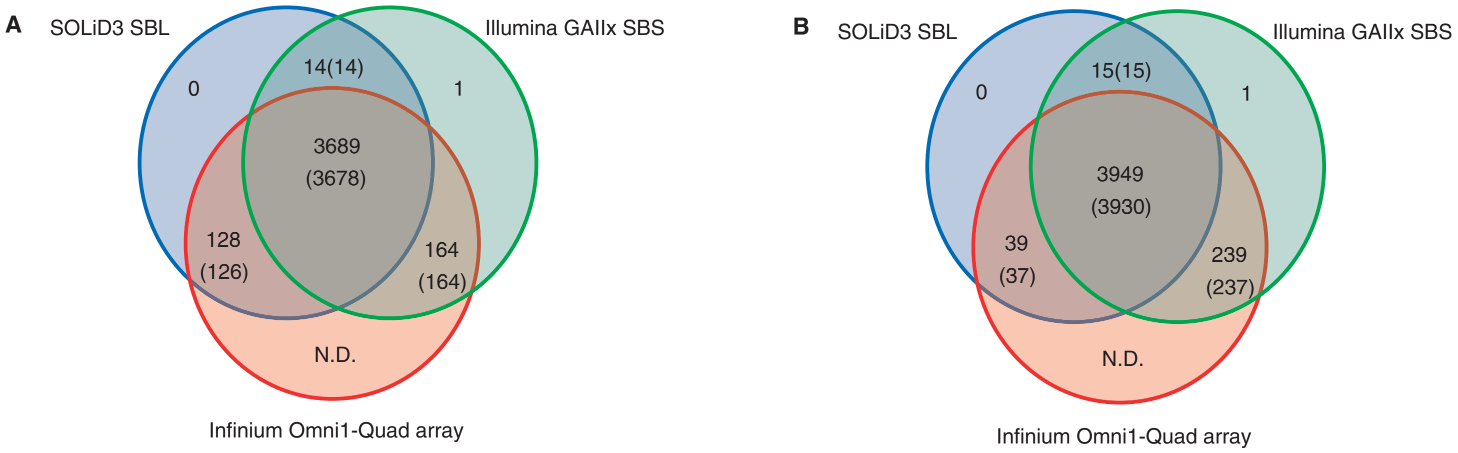
Target Coverage



Median coverage increased asymptotically with sequence depth

~2.6 Gb of sequence is necessary for coverage and accuracy

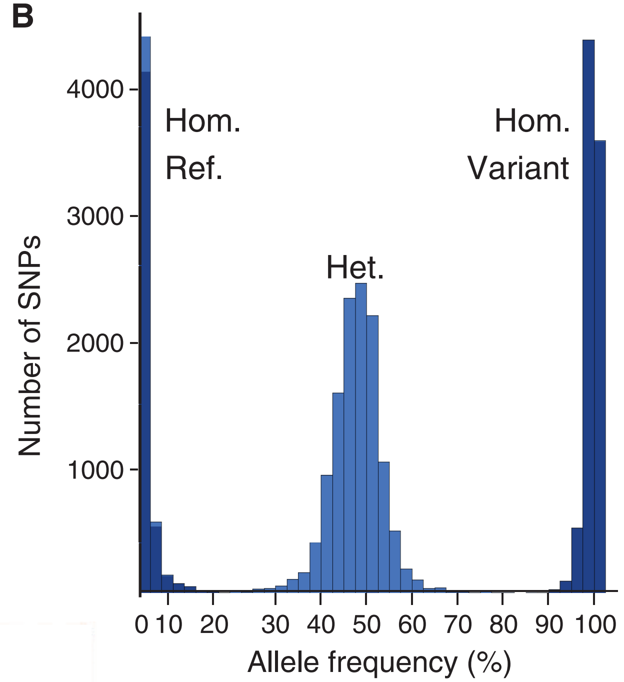
NGS on SNP calls



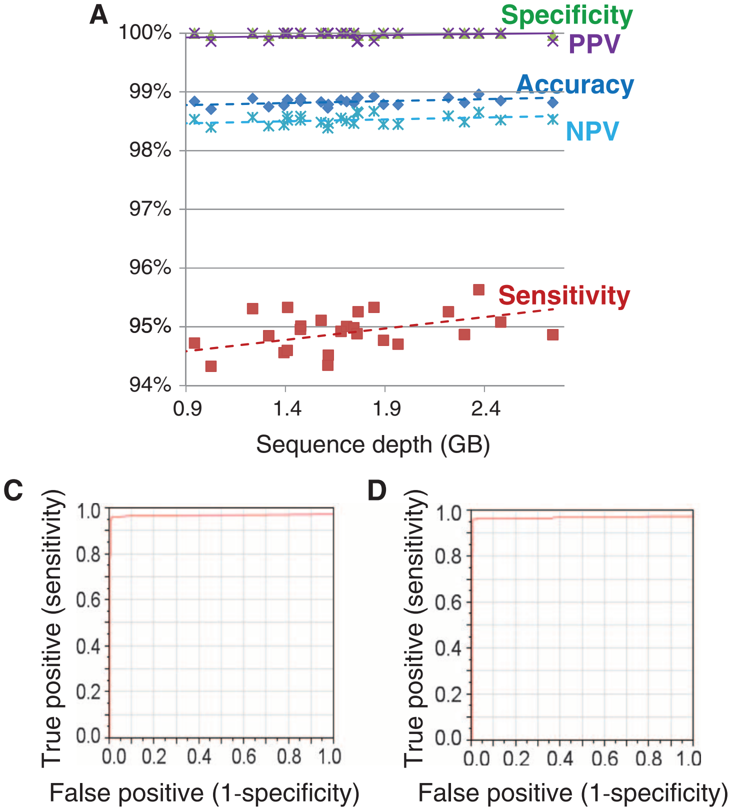
SNPs are called if present in > 10 uniquely aligning reads (left figure) or > 4 uniquely aligning reads (right figure), with average quality score > 20.

SNP Genotype and Accuracy

Distribution of read count-based allele frequency



92,106 SNP calls in 26 samples



Accuracy of SNP genotyping against Infinium array

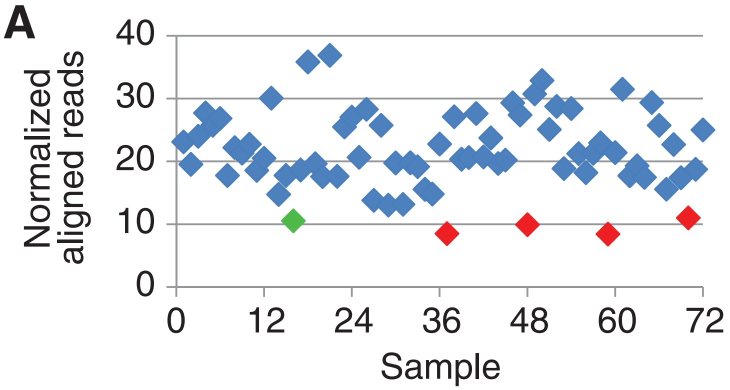
Detect Gross Deletion

Use HiSeq NGS method

Reduce penalty on polynucleotide variants [-1 - log(indel-length)]

Detect gross deletion by perfect alignment to mutant junction reference sequences or by local decrease in normalized coverage.

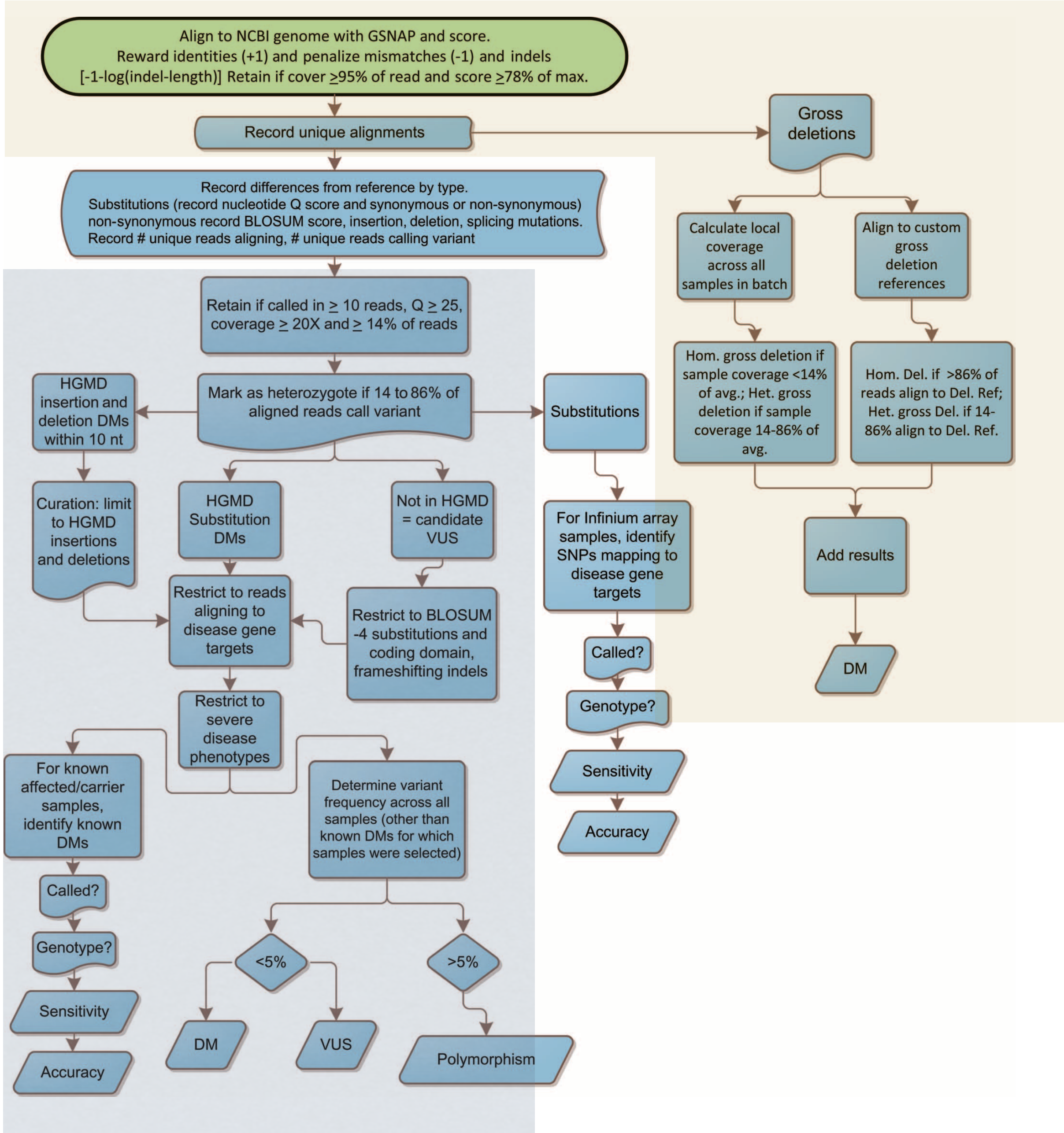
Deletion in CLN3



Four known heterozygotes (red) and one undescribed carrier (green) are identified

Reads were normalized to total sequence generated in a batch

Decision Tree



Align to NCBI genome with GSNAP and score. Reward identities (+1) and penalize mismatches (-1) and indels [-1-log(indel-length)] Retain if cover ≥95% of read and score ≥78% of max.

Record unique alignments

Gross deletions

Calculate local coverage across all samples in batch

Align to custom gross deletion references

Hom. gross deletion if sample coverage <14% of avg.; Het. gross deletion if sample coverage 14-86% of avg.

Hom. Del. if >86% of reads align to Del. Ref; Het. gross Del. if 14-86% align to Del. Ref.

Add results

DM

Record differences from reference by type. Substitutions (record nucleotide Q score and synonymous or non-synonymous) non-synonymous record BLOSUM score, insertion, deletion, splicing mutations. Record # unique reads aligning, # unique reads calling variant

Retain if called in ≥ 10 reads, Q ≥ 25, coverage ≥ 20X and ≥ 14% of reads

Mark as heterozygote if 14 to 86% of aligned reads call variant

HGMD insertion and deletion DMs within 10 nt

Curation: limit to HGMD insertions and deletions

HGMD Substitution DMs

Not in HGMD = candidate VUS

Restrict to reads aligning to disease gene targets

Restrict to BLOSUM -4 substitutions and coding domain, frameshifting indels

Restrict to severe disease phenotypes

For known affected/carrier samples, identify known DMs

Called?

Genotype?

Sensitivity

Accuracy

Determine variant frequency across all samples (other than known DMs for which samples were selected)

<5%

>5%

DM

VUS

Polymorphism

Substitutions

For Infinium array samples, identify SNPs mapping to disease gene targets

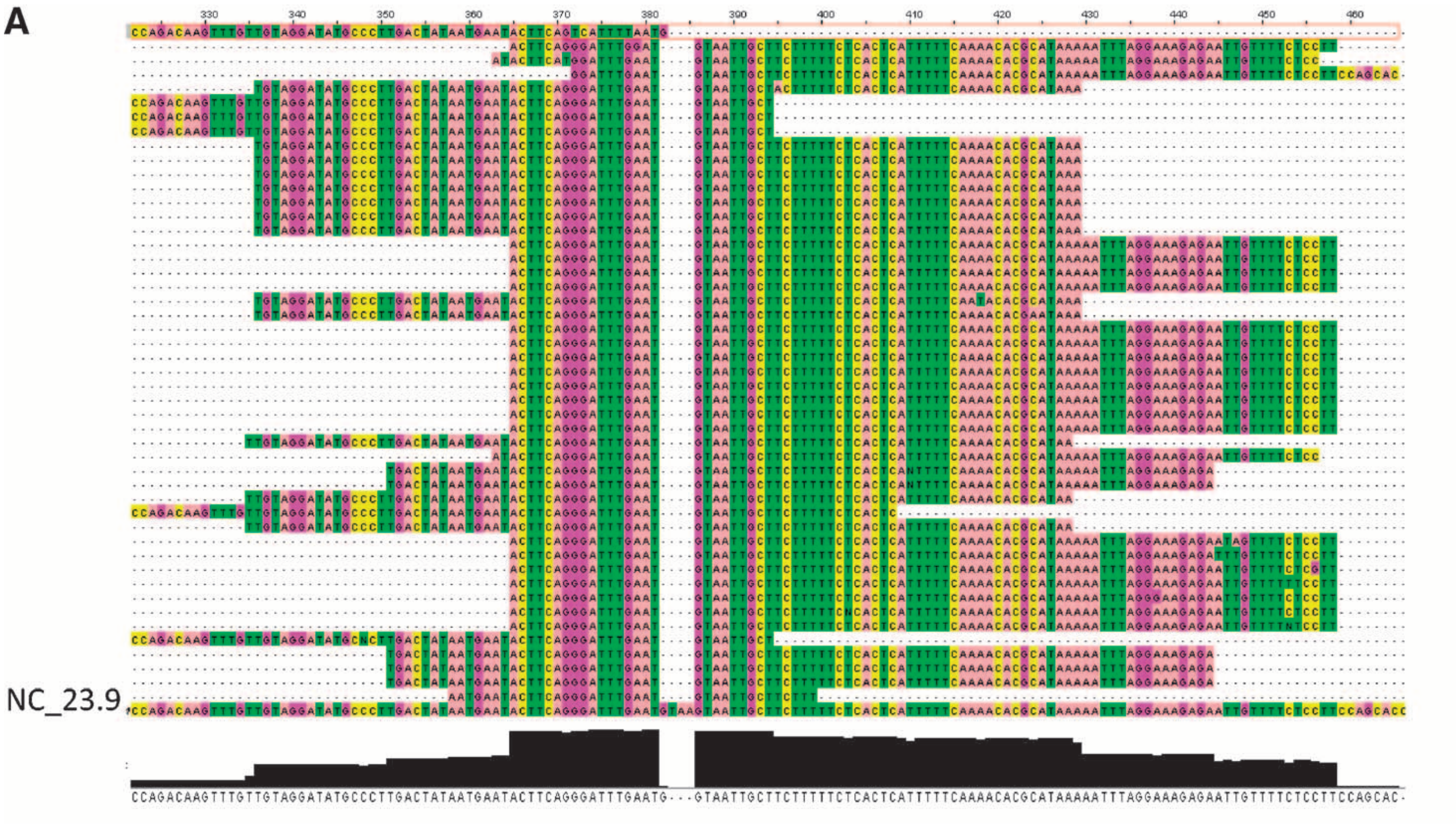
Called?

Genotype?

Sensitivity

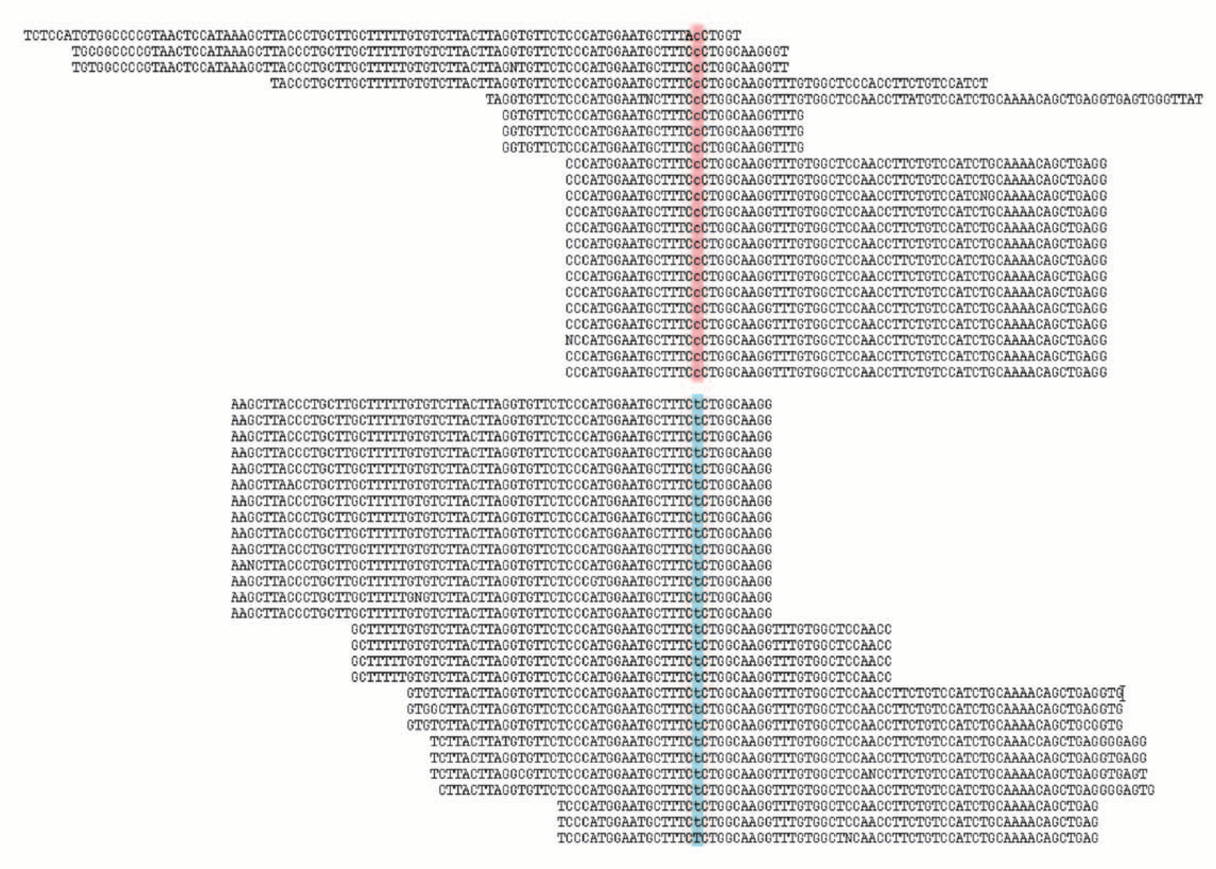
Accuracy

Some Incorrect Annotations


A
330
340
350
360
370
380
390
400
410
420
430
440
450
460
NC_23.9
CCAGACAAGTTTGTTGTAGGATATGCCCTTGACTATAATGAATACTTCAGGGATTTGAATG---GTAATTGCTTCTTTTTCTCACTCATTTTTCAAAACACGCATAAAAATTTAGGAAAGAGAATTGTTTTCTCCTTCCAGCAC-

Sample: an affected male with X-linked recessive Lesch-Nyhan syndrome
Before: characterized as deletion of HPRT1 exon 8 by cDNA sequencing
Actual: splicing mutation of IVS intron 8.

Some Incorrect Annotatios



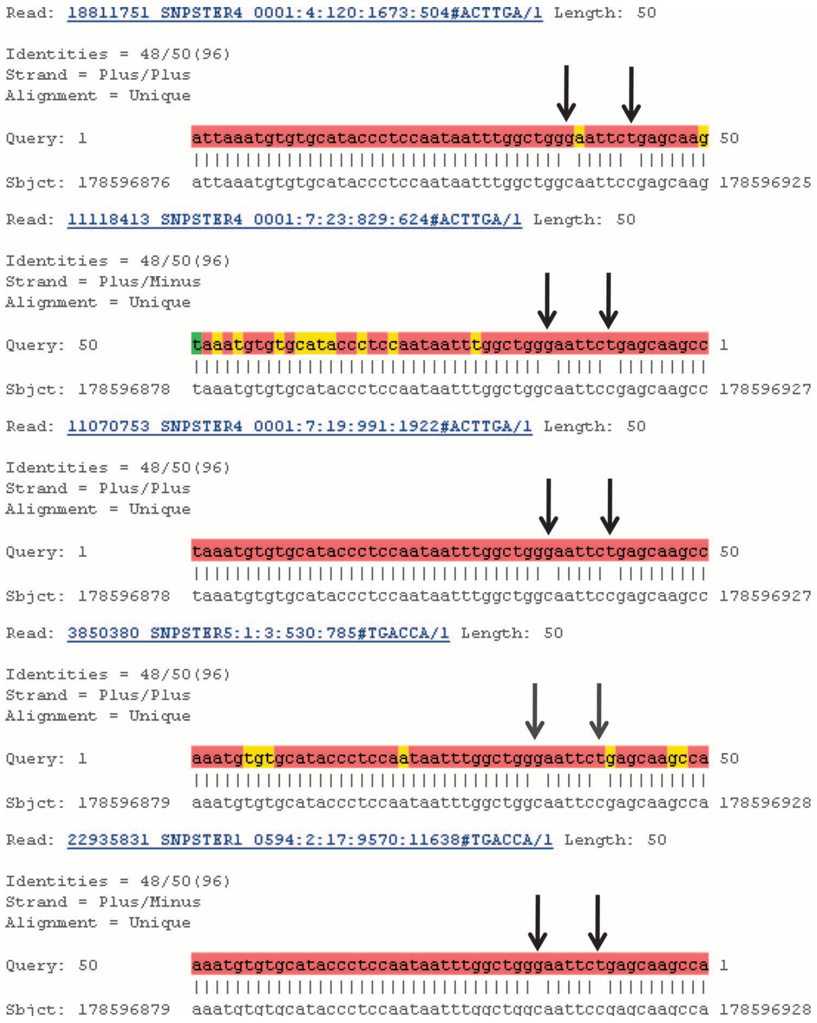
```
TCTCCATGTGGCCCCGTAACTCCATAAAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTAcCTGGT
      TGCGGCCCCGTAACTCCATAAAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGGT
      TGTGGCCCCGTAACTCCATAAAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGNTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGTT
                               TACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCCACCTTCTGTCCATCT
                                                          TAGGTGTTCTCCCATGGAATNCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTATGTCCATCTGCAAAACAGCTGAGGTGAGTGGGTTAT
                                                            GGTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGTTTG
                                                            GGTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGTTTG
                                                            GGTGTTCTCCCATGGAATGCTTTCcCTGGCAAGGTTTG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCNGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    NCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG
                                                                    CCCATGGAATGCTTTCcCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGG

                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTAACCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTAGCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AANCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCGTGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGNGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                          AAGCTTACCCTGCTTGCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGG
                                         GCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACC
                                         GCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACC
                                         GCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACC
                                         GCTTTTTGTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACC
                                                GTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGGTG
                                                GTGGCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGGTG
                                                GTGTCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGCGGTG
                                                   TCTTACTTATGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAACCAGCTGAGGGGAGG
                                                   TCTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGGTGAGG
                                                   TCTTACTTAGGCGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCANCCTTCTGTCCATCTGCAAAACAGCTGAGGTGAGT
                                                    CTTACTTAGGTGTTCTCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAGGGGAGTG
                                                                   TCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAG
                                                                   TCCCATGGAATGCTTTCtCTGGCAAGGTTTGTGGCTCCAACCTTCTGTCCATCTGCAAAACAGCTGAG
                                                                   TCCCATGGAATGCTTTCTCTGGCAAGGTTTGTGGCTNCAACCTTCTGTCCATCTGCAAAACAGCTGAG
```

Sample: an affected male with X-linked recessive Pelizaeus-Merzbacher disease

Before: substitution mutation in PLP1 exon 5 c.67C>T, P215S

Actual: PLP1 gene duplication

Some Incorrect Annotatios



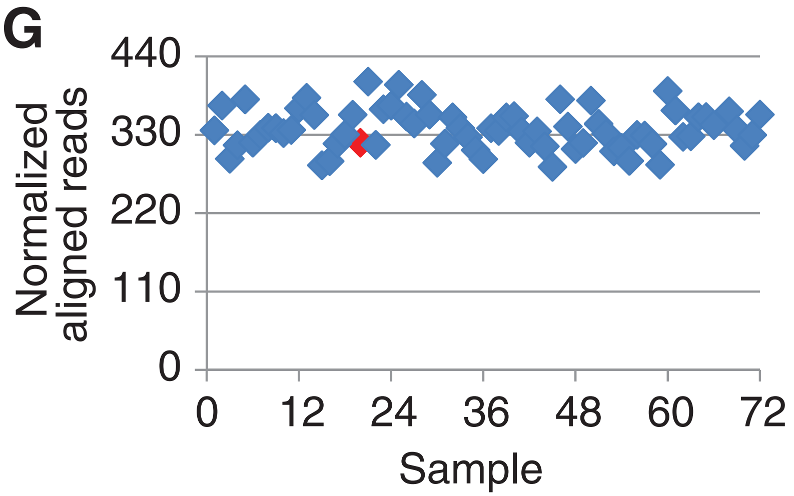
```
Read: 18811751 SNPSTER4 0001:4:120:1673:504#ACTTGA/1 Length: 50

Identities = 48/50(96)
Strand = Plus/Plus
Alignment = Unique

Query: 1          attaaatgtgtgcataccctccaataatttggctgggaattctgagcaag 50
                  |||||||||||||||||||||||||||||||||||| ||||| |||||||
Sbjct: 178596876  attaaatgtgtgcataccctccaataatttggctggcaattccgagcaag 178596925

Read: 11118413 SNPSTER4 0001:7:23:829:624#ACTTGA/1 Length: 50

Identities = 48/50(96)
Strand = Plus/Minus
Alignment = Unique

Query: 50         taaatgtgtgcataccctccaataatttggctgggaattctgagcaagcc 1
                  |||||||||||||||||||||||||||||||||| ||||| |||||||||
Sbjct: 178596878  taaatgtgtgcataccctccaataatttggctggcaattccgagcaagcc 178596927

Read: 11070753 SNPSTER4 0001:7:19:991:1922#ACTTGA/1 Length: 50

Identities = 48/50(96)
Strand = Plus/Plus
Alignment = Unique

Query: 1          taaatgtgtgcataccctccaataatttggctgggaattctgagcaagcc 50
                  |||||||||||||||||||||||||||||||||| ||||| |||||||||
Sbjct: 178596878  taaatgtgtgcataccctccaataatttggctggcaattccgagcaagcc 178596927

Read: 3850380 SNPSTER5:1:3:530:785#TGACCA/1 Length: 50

Identities = 48/50(96)
Strand = Plus/Plus
Alignment = Unique

Query: 1          aaatgtgtgcataccctccaataatttggctgggaattctgagcaagcca 50
                  ||||||||||||||||||||||||||||||||| ||||| ||||||||||
Sbjct: 178596879  aaatgtgtgcataccctccaataatttggctggcaattccgagcaagcca 178596928

Read: 22935831 SNPSTER1 0594:2:17:9570:11638#TGACCA/1 Length: 50

Identities = 48/50(96)
Strand = Plus/Minus
Alignment = Unique

Query: 50         aaatgtgtgcataccctccaataatttggctgggaattctgagcaagcca 1
                  ||||||||||||||||||||||||||||||||| ||||| ||||||||||
Sbjct: 178596879  aaatgtgtgcataccctccaataatttggctggcaattccgagcaagcca 178596928
```

Sample: an affected female with aspartylgucosaminuria

Before: characterized as compound heterozygotes

Actual: homozygous for two adjacent substitutions

Some Incorrect Annotatios

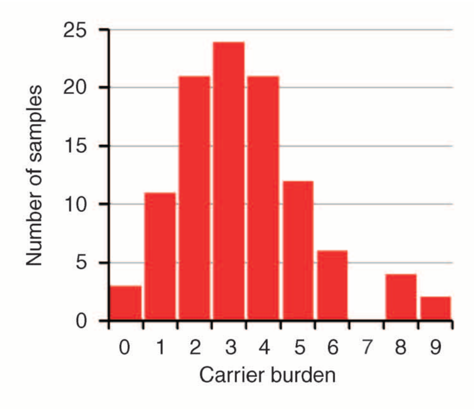


Sample: affected with Cockayne syndrome B

Before: deletion of ERCC6 exon 9

Actual: no gross deletion was observed

Carrier Burden



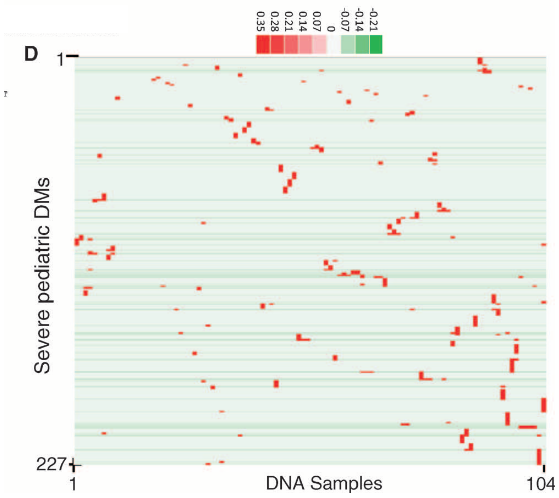
336 variants were retained as likely disease mutations in 104 samples;

A variant was retained reported in HGMD and literature, had been shown to result in LOF, was the only variant in affected individuals and absent in control, and was predicted to result in premature stop codon or loss substantial protein portion

Average: 2.8 / genome

Ward hierarchical clustering of 227 DM in 104 samples

Resulting pattern is random, suggesting that targeted population testing is likely to be ineffective



Conclusions

- Described a screening test (target capture + NGS) for carriers of 448 severe childhood recessive diseases
- Found a list of incorrect literature-annotated disease mutations
- Estimated the average carrier burden (2.8) of disease mutations causing severe childhood recessive diseases.

Future Challenges

- Refinement of list of diseases
- Automation, software implementation
- Validation in realistic testing situations featuring investigator blinding
- Ethic concerns