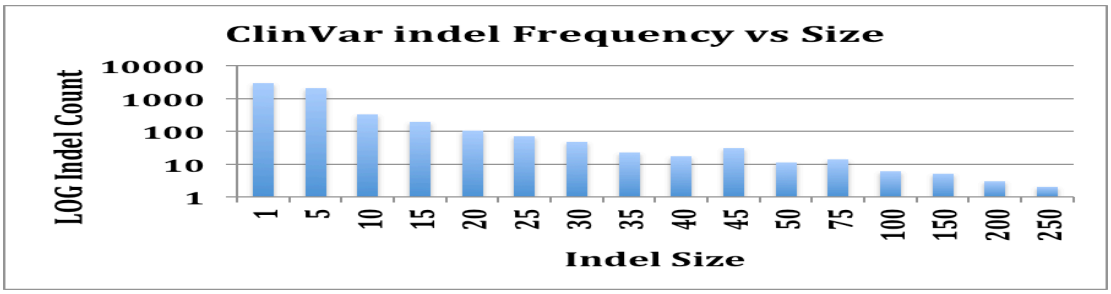
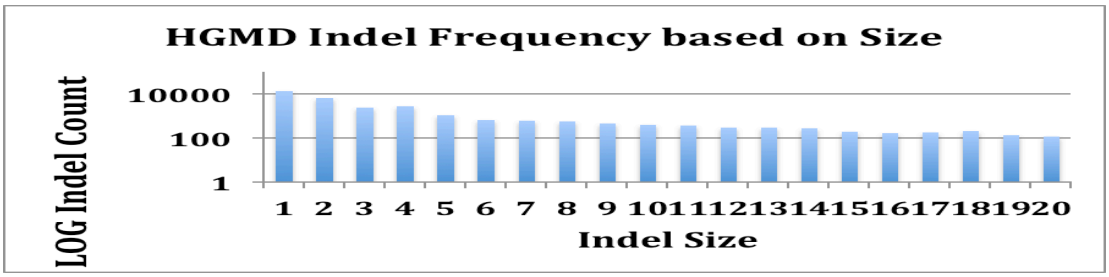
Expanding the Utility of Next Generation Sequencing for Unified Genetic Testing

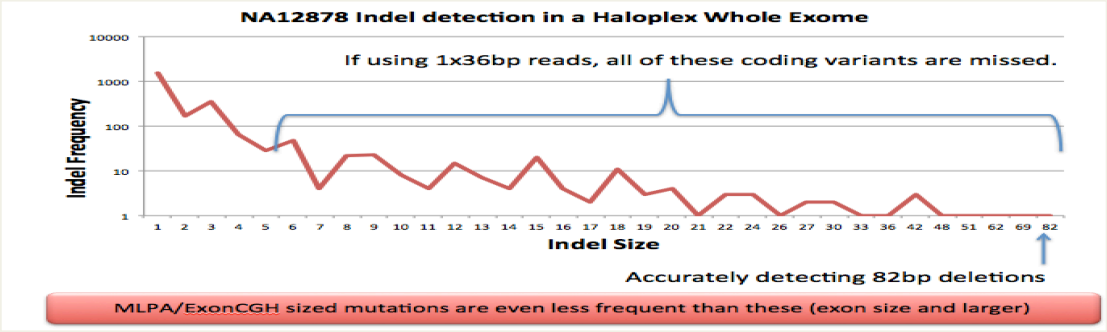
Courtagen is committed to helping physicians, families, and children suffering with neurological disorders. Our test results provide doctors with genetic information to help diagnose disease and guide treatment decisions. Genetic testing can lead to positive changes or, in some cases, major improvements in a child's quality of life.

Currently genetic lesions are tested with a multitude of technologies. Single nucleotide polymorphisms are easily detected with short read next generation sequencers. Exon sized deletions (del-dups) are traditionally detected with MLPA or exon CGH. Larger sized copy number alterations (500kb+) are detected with chromosomal microarrays (CMA). Even with three to four technologies in use, blind spots remain in clinical sensitivity. In particular, insertions and deletions in the 5bp-150bp size range have been difficult to detect with any technology. Sanger sequencing often captures these, but does not scale to gene panels or exome sized sequencing capacity.



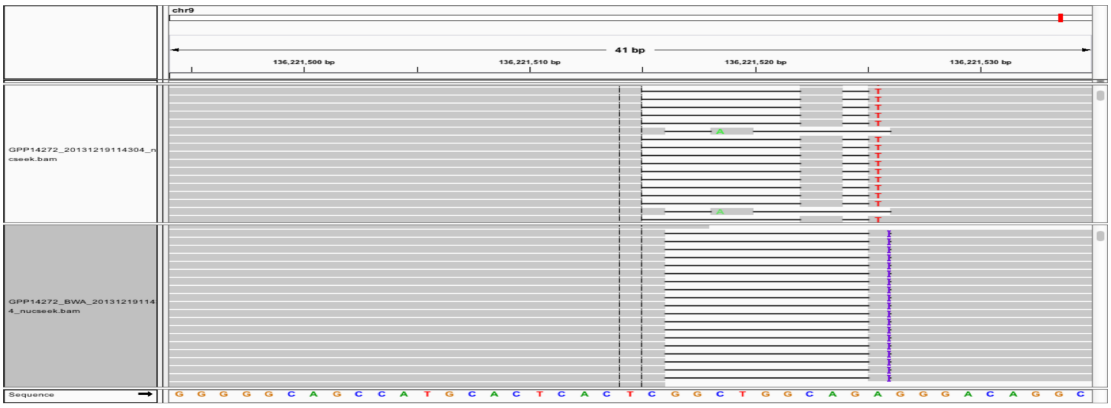
This is a clinically relevant blind spot as 10 fold more variants exist in ClinVar and HGMD in this size range than all of the clinical variants larger than 150bp. Most whole genome sequencing projects support this mutational distribution in the human genome, and due to the described sequencing technical limitations, it's quite likely variants in this size range are under-reported in the clinical databases.[1-3] Removing this blind spot will enable clinical sensitivity of more deleterious lesions than what can be detected by both MLPA and CMA. This testing process could be streamlined if one technology could detect the majority of the lesions and with the increased read lengths of next generation sequencers, detection of these lesions are within reach. With the improving read lengths of next generation sequencer like the Ilumina MiSeq, 250bp to 600bp read lengths are now achievable and this enables better indel detection.





Improved Mapping Algorithms

Preliminary sequencing with 2x250bp sequencing utilizing Agilents Haloplex 37Mb Exome capture system on NIST sample NA 12878 has revealed detection of 50bp and greater deletions. These longer readlengths require new mapping tools to fully leverage their utility. Conventional mapping tools were trained on short reads, and often fail to make use of the readlength achievable in 2014. Below is a comparison of Bowtie2 vs BWA-MEM properly aligning a large indel. The Bowtie2 (top) alignment genotypes this indel as 3 different events. The BWA-MEM (bottom) alignment makes the correct 9bp indel call while the Bowtie2 calls this a 6bp and a single base indel with a penultimate SNV.



Improved Target Capture

In addition to longer reads and improved read mapping tools, hybridization based capture protocols need to be re-evaluated. Hybridization based capture protocols generate 30-40% of their reads as off target mappings. These off target mappings are the result of low stringency capture, and can introduce many false positive indels by capturing and collapsing pseudogenes on the targets of interest. Polymerase or ligation based protocols such as multiplex PCR or Haloplex introduce a high fidelity enzyme that interrogates both ends of the molecule being captured. This high stringency provides over 90% of the reads being on target but introduces other biases that need to be compensated for. This stringent capture can suffer from allelic dropout due to primer sites targeting regions of the patient's genome that have private variants. As a result hyper-tiling regions of interest is required to eliminate this potential for allelic drop out. Tiling a target 10X is preferred.

Allele Drop Out

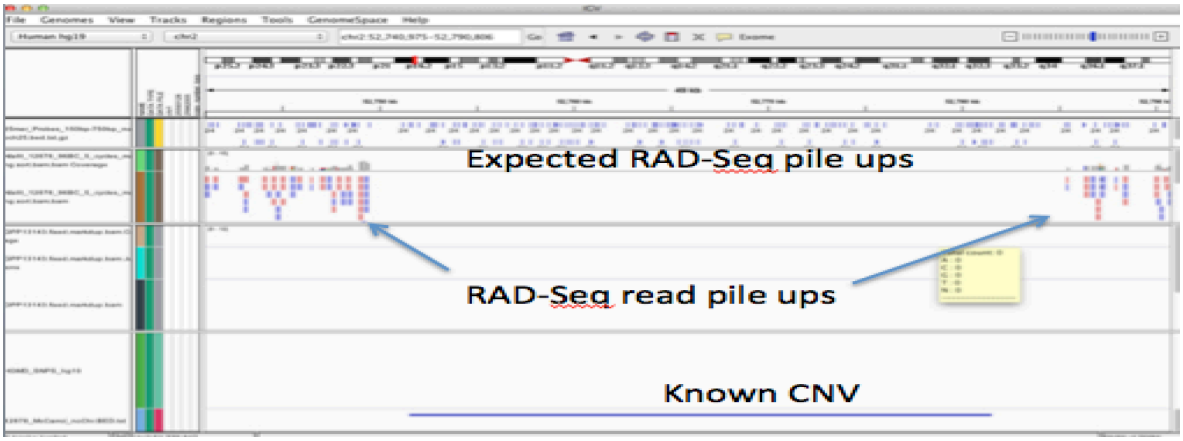
Courtagen DIAGNOSTICS LABORATORY

Allele drop out is a function of the number of amplicons per target.
The fewer the number, the more prone the read out is to primer sequences amplifying the allele which matches your reference used in oligo design.

- HaloPlex™, custom DNA capture: on average has 10 amplicons tiling every base.
- Other methods: only 1-2 amplicons per target subjecting amplification drop out due to unlucky primer design.

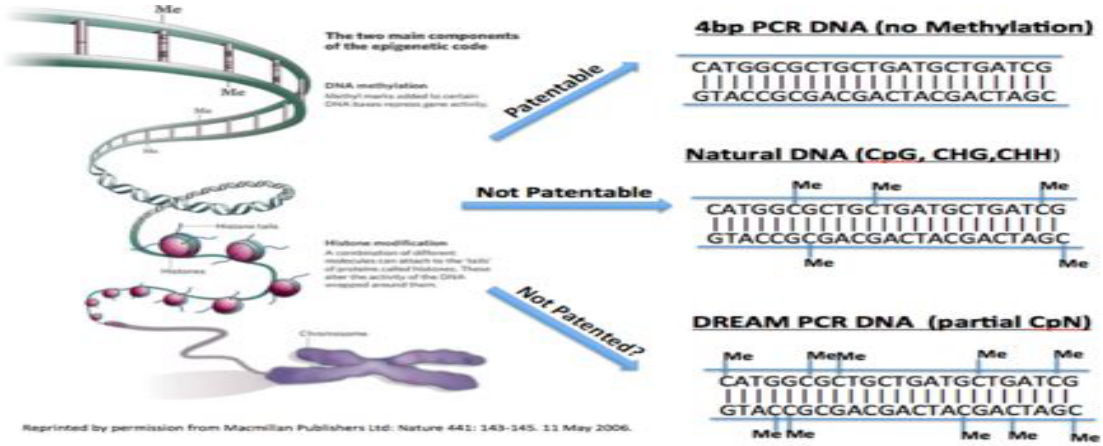
CMA to RAD-Seq

To further close the blind spots in sequencing, Courtagen has developed a RAD-Seq assay to replace CMA. Zheng *et al*. demonstrate how this SOLiD sequencing based RAD-Seq assay can out perform a 1.8 million Affymetrix SNP array for detecting copy number variants[4]. Modifications to this assay have enabled 100ng CMA to be performed from saliva in less than 48hrs at Courtagen. Sensitivity and specificity is still being tested with well-characterized Coriell samples, but preliminary data is supportive of Zheng *et al.* findings. The depth of the pile-ups in RAD-Seq also enables LOH detection to be performed equal or greater resolution. Most CMA today is performed on Agilent 180K arrays, thus RAD-Seq is expected to deliver higher resolution and higher quality data and this expectation is well described in *Zheng et al.* publication.



Gene Patents Can Create False Negatives

The other source of false negatives in sequencing epileptic patients is due to simple omission of genes due to intellectual property. This issue continues to plague the US market as the SCOTUS decision in 2013 reinforced Myriad's position with companies using sequencing techniques which man modify the DNA with Next Generation Sequencers or target capture. Courtagen has invested significantly in the design of methods that capture DNA such that it is indistinguishable from native DNA. Man modifications that fail to be substantively different than natural DNA are not patentable. This was published in Nature Biotechnology in October of 2013[5]. Built into the design of this sequencing technique is a PCR decontamination tool very helpful for patient privacy and heteroplasmy detection.



The impact on Epilepsy

Closing these blind spots in clinical sequencing is of particular interest in Dravet syndrome as 10-15% of these patients have mutations in SCN1A that are larger than a single nucleotide variant. This drug resistant form of epilepsy has recently shown new therapeutic promise with cannabidiol (CBD). Preliminary response to CBD is promising, but non-responders exist and better understanding of the genetics of Dravet syndrome may stratify this response. This underscores the importance of resolving clinical blind spots in SCN1A testing. GW Pharmaceuticals is currently funding a study that provides CBD to 125 patients, and Courtagen has volunteered to sequence these patients. This cohort is only 40% Dravet syndrome and thus very few patients with lesions in this size range are expected.

Specific Aims

Courtagen is reaching out to various epilepsy organizations to obtain Dravet samples. Sequencing is being performed and returned to collaborating partners given prior genetic profiling has already been performed and we can benchmark our test. We are most interested in samples that are known to have deldupes or have been diagnosed with Dravet syndrome and their sequencing failed to find the putative lesion.

epiSEEK®: Comprehensive Sequence Analysis for Epilepsy and Seizure Disorders

Epilepsy is a common neurological disorder with a highly heterogeneous background and a strong genetic contribution. Courtagen Diagnostics Labs' epiSEEK™ test provides the complete sequence of 327 genes that have been reported in association with epileptic phenotypes. The gene categories are listed below.

Gene Categories:
• Glycosylation Disorders (23 genes)
• Early Infantile Encephalopathy (16 genes)
• Idiopathic Generalized Epilepsy (28 genes)
• Mitochondrial Dysfunction (27 genes)
• Inherited Metabolic Diseases (41 genes)
• Neurodegeneration (32 genes)
• Joubert Syndrome (9 genes)
• Epilepsy in X-linked Mental Retardation (18 genes)
• Brain or Nervous System Malformations (50 genes)
• Syndromic Disorders (71 genes)
• Other (12 genes)

Assay Features

- Variant Confirmation – Sanger validation confirms accuracy of clinically relevant variants run on our next generation sequencing assay
- Proprietary non hybridization capture technique to avoid pseudogenes
- Our proprietary ZiPhyr™ Bioinformatics pipeline leverages standard industry bioinformatics engines in conjunction with our unique analysis workflows to generate actionable results for our customers.
- Clinical Interpretation by our medical officer, Dr. Richard Boles
- Data Delivery – Web portal for ordering, reviewing and downloading clinical results. Clinical reports on each patient are also available via secure fax or mail
- Quality- For variant calling we are using Phred scores of at least 100, but typical values are in the 1000s. There is less than 1 chance in 10^10 that this is an error.

  The Phred scaled probability of probability that REF/ALT polymorphism exists at this site given sequencing data. Because the Phred scale is -10 * log(1-p), a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in 10^10 chance. These values can grow very large when a large amount of NGS data is used for variant calling.

Data Delivery:

- VCF files or raw read files via disk drive ($300 per drive shipped)
- Web portal data download: Free of Charge
- Fax or Mail of physical clinical report: Free if Charge
- Email reports only if patient data is blinded to comply with GINA/HIPAA

Required and desired amounts of DNA: Desired DNA: 1µg DNA at >20ng/µl
We have designed everything to work with pediatric saliva samples. We have also validated, blood and tissue. A single test requires 200ng of DNA. We ask for 1µg to account for re-work and variance in quantification tools in the field. In addition we utilize this stock for Sanger confirmation of clinically significant variants.

How do you require us to prep the DNA for you?
We can prep the DNA if you have saliva or blood. Alternatively, Beckman SPRI-TE genomic DNA purifications are used internally.

DNA concentration required? Greater than 20ng/µl

What are your laboratory qualifications and/or certificate numbers?
We are CLIA certified facility #5313, License # 3163, Classification: Full. Our CMS CLIA ID Number is 22D2035168. Our firm has also CAP accredited #8044831. Additional livensing information can be found at http://www.courtagen.com/company-licensing.htm

Which NGS Platform is used?
We use Paired 250bp reads on the Illumina MiSeq. Molecular Inversion Probes and PCR techniques are utilized to enrich ~ 5,000 exons. Mean Sequencing depth is >200X. We will not ship any data unless 95% of the targets are sequenced 20X or higher.
SNP are only called on consensus Phred scores over Phred 100.

How is each variant confirmed?
We may sanger sequence variants of clinical significance for confirmation. Significance is measured with 5 algorithms:

- Evolutionary Conservation scores across, Primates, Mammals, Birds, Reptiles,
- Marsupials, Fish, and invertebrates.
- Mutation Taster
- SIFT
- PolyPhen2
- Mutationassessor

How will prior rare variants determined in other studies be used as a check against putative rare variants in our samples?
All variants are screened against 1,000 genomes data, dbSNP,HGMD, OMIM and our own internal database.

Courtagen clinical report in electronic or paper format. The following delivery formats are available:

- Web portal
- Fax or mail
- Disk drives can include VCF and raw data if requested
- Email only if patient data is blinded to comply with HIPAA/GINA regulations
- An example report is attached
- Two abstracts are attached which describe similar MiSeq Courtagen assays.