Dovetail GENOMICS

TECH SHEET

Dynamic Genome Profiling

with Fix-C™,Dovetail™ Hi-C Kit and Selva™ SV Detection

Accurately detecting and resolving large and complex structural variants (SVs) requires genomic information that spans the breakpoints and accurately aligning them within the reference genome. This has been a challenge until now.

FEATURES

- Dovetail™ proximity ligation captures megabase-range information needed for detecting and resolving the largest and most complex variants
- Fix-C™ captures megabase-range information from FFPE, previously not considered to contain long-range information
- Selva™ analysis detects the largest and most complex structural variants with high confidence and includes tools for visual investigation
- A customer focused service business with a proven record of delivering proximity-ligation solutions to genomic challenges

Dovetail™ Proximity Ligation Spans Megabases Needed to Discover and Resolve

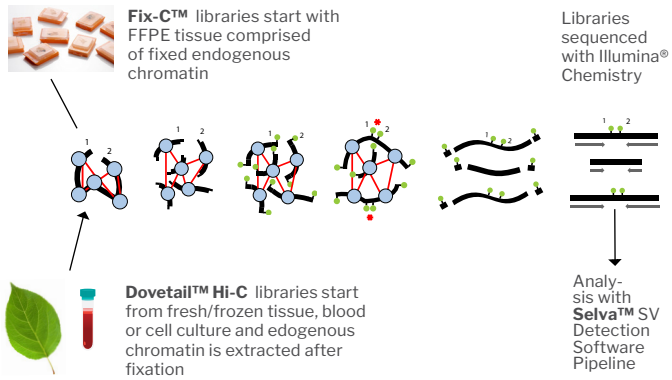
Structural variants (SVs) account for more base pair differences (60%) than single nucleotide variants and smalls indels combined (40%) and are of increasing interest for their roles in diseases from cancer to schizophrenia.

Even with an increased focus, accurately detecting and resolving large SVs on a genome wide scale has remained a challenge, with many having multiple breakpoints, repetitive regions overlapping the breakpoints, or distances between breakpoints that exceed the insert sizes of the sequencing technology. Technologies are increasingly relying on longer reads and more complex bioinformatics pipelines with inherent limitations and drawbacks like the length of the input DNA and decreased confidence of calls.

Dovetail solves these challenges by harnessing the power of proximity ligation technology. This provides information on genomic connections ranging from hundreds of bps to Mbps, ideal for resolving even the largest and most complex structural variants.

Selva™ detects balanced, unbalanced, and complex events with no upper limit on the size of the variant. It works by identifying linkages between rearranged regions from bps to megabases apart generated by Dovetail's proximity ligation techniques. SV calls from Selva™ and the underlying data can be further investigated with an interactive viewer that displays all SV calls, gene tracks, and has features for searching, multiple sample analysis, panning and zooming.

Figure 1. Proximity ligation workflow. The Fix-C™ technique starts with FFPE and Dovetail™ Hi-C technique starts with fresh samples, both generating molecules spanning hundreds of bps to Mbps.



Fix-C™ Reconstructs FFPE DNA to Generate Long-Range Information and Find Large SVs

The Dovetail™ Hi-C technique is able to span megabases in fresh samples with DNA in the form of intact endogenous chromatin. The condition of DNA in FFPE samples is more challenging.

The process for preserving samples entails crosslinking the DNA with formalin which breaks the DNA and greatly reduces the insert sizes. This has made it impossible to detect and resolve large SVs from FFPE samples with available sequencing technologies.

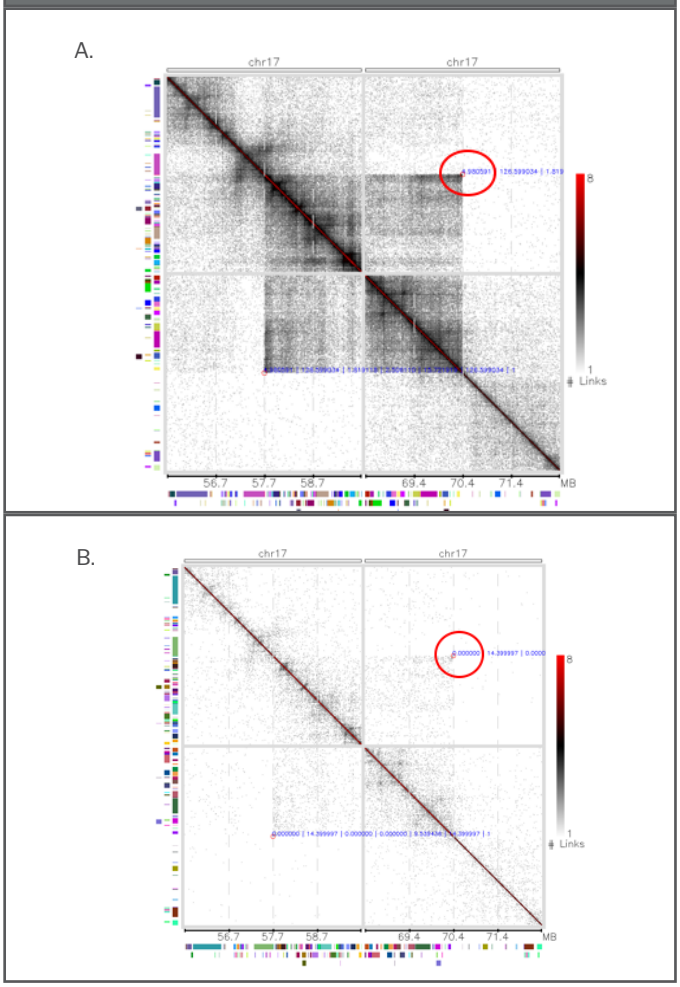
Interestingly, the cross-linking that breaks the DNA also captures the proximity relationships within the chromatin. These relationships are leveraged by the Fix-C™ technique to span megabases of genomic space on the scale with un-fixed material. To compare the performance of fixed and un-fixed tissues, a study was conducted with 10 matched fresh tissue and FFPE breast cancer tumor samples. The samples were processed with Dovetail™ Hi-C and Fix-C™ respectively, sequenced to 200M read-pairs in 2X150 PE format, and analyzed with Selva™ software. The study showed 100% concordance, with the same large rearrangements detected in the fresh tissue and matched FFPE tissue. No other long-range sequencing technologies claim to detect large SVs from FFPE.

Selva™ Is Validated to Detect the Largest and Most Complex Structural Variants with High Confidence

To validate Dovetail's techniques and Selva™ software, a study was conducted with 12 FFPE tumor samples ranging from 20% to 90% tumor with FISH results. The samples were processed with Fix-C™, sequenced to 200M read-pairs in a 2X150 PE format, and analyzed with Selva™.

The software called 9 of the 10 FISH-detected variants, and upon further validation, the negative Selva™ call proved to be a false positive FISH call.

Figure 2. Contact map comparisons of matched FFPE and fresh breast cancer tumor sample. Shows the same intra-chromosomal structural variant in chromosome 17, both called by the Selva™ pipeline. A. Flash Frozen breast tissue, B. FFPE breast tumor



See ordering details on page 4 to place an order or for more information.
visit us at www.dovetailgenomics.com or send an email to info@dovetail-genomics.com

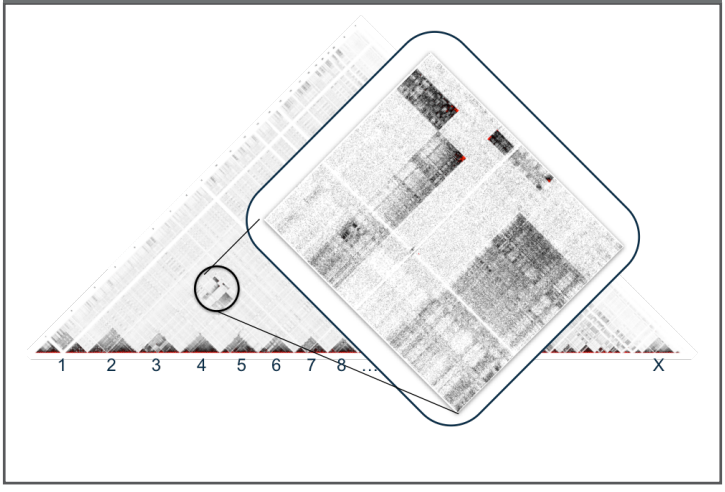
Table 2. Summary of samples tested, FISH and Selva™ fusion detection, and orthogonal findings.

Sample Number	Histology	Tumor Percentage	FISH Fusion Call	Selva™ Call	Additional Othogonal Validation Findings	Total Selva™ Calls
1	Lung ardenocarcinoma	20	ALK+	ALK (-)	validated false positive FISH call	1
2	Adenoid cystic carcinoma	50	MYB+	EWSR1-MYB	validated	10
3	Round cell liopsarcoma	90	FUS+	DDIT3-FUS	validated	1
4	Extraskeletal myxoid chondrosarcoma	60	EWSR1+	EWSR1-NR4A3	validated	2
5	Synovial sarcoma	90	SS18+	PAOX-SS18 SS18-22X2B	validated	9
6	Mammary analog secretory carcinoma	30	ETV6+	ETV6-NTRK3	validated	2
7	Lung adenocarcinoma	50	ROS (-)	MYOSC-ROS1	validated false negative FISH call	11
8	Lung adenocarcinoma	60	RET (-)	RET (-)	validated	0
9	Angiomatoid fibrous histiocytoma	30	EWSR1+	EWSR1-CREB1	validated	3
10	Inflammatory myofibroblastic tumor	20	ALK+	CLTC-ALK	validated	1
11	Adenoid cystic carcinoma	80	MYB+	MYB-NFIB	validated	4
12	Synovial sarcoma	80	SS18+	SS18-SSX2B	FISH validated only	2

In addition, Selva™ called a potential clinically relevant ROS fusion, MYO5C-ROS1, missed by FISH and validated by orthogonal technologies.

As expected, Selva™ also detected additional large variants (N=36) across the 12 samples not assayed for by FISH that have either been validated by orthogonal technologies or meet reporting criteria. Several of the variants called by Selva™ were highly complex and could only be resolved with Dovetail proximity ligation technology as shown in Figure 3. As a measure of specificity, regions with Selva™ calls validated by FISH were compared to the same regions in samples with different FISH calls serving as a specificity control. There were expected proximity signals and detected calls in all regions of samples with validated FISH calls and no observed signals above background or detected calls in the same regions for samples with FISH calls in other regions. This indicates the validated proximity signals called by Selva™ are specific for SVs.

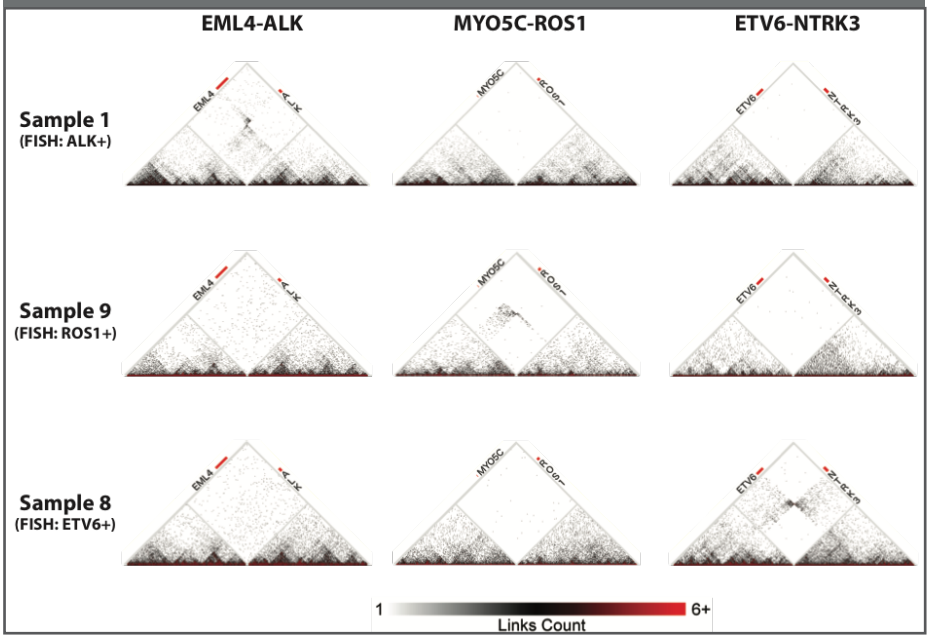
Figure 3: Selva™ detection of known and novel large-scale structural variation in clinical FFPE samples. Fix-C™ with Selva™ detection of a FISH-confirmed EWSR1+ fusion and a complex gross rearrangement between chromosomes 3 and 6.



Currently, there is a lack of gold standard data sets containing a wide range of validated SVs larger than 50kb. To further benchmark Selva™, balanced and unbalanced SVs of varying sizes with breakpoints covering all classes were generated with synthetic breaks in real Hi-C data sets with 200M raw read-pairs.

For balanced SVs between 50kb and 10Mb in size with 100% variant frequency, Selva's™ sensitivity was 96%. For unbalanced SVs over the same range, the sensitivity was 85%. Selva performed down to 20% variant frequency with a sensitivity of 67% for balanced SVs and 65% for unbalanced SVs ranging from 50kb to 10Mb in size. For comparison, long read sequencing technologies only claim confident detection of SVs up to 50kb in size, and have no claims for SV detection from FFPE tissue. Selva confidently detects the larger SVs missed by current sequencing methods from fresh tissue and FFPE.

Figure 4. Fix-C™ with Selva™ detection of genomic rearrangements in clinical samples. FISH-confirmed ALK+ (sample 1), ROS1+ (sample 9) and ETV6+ (sample 8) gene fusion events are detected by Fix-C. Samples known to harbor genomic rearrangements show strong signal of proximity between the examined loci while others act as controls, displaying only background signal between the same loci.



ORDERING INFORMATION

Contact Orders@dovetail-genomics.com or (831) 713-44656