Redefining nanopore sequencing with Duplex Sequencing by Expansion (SBX-D)

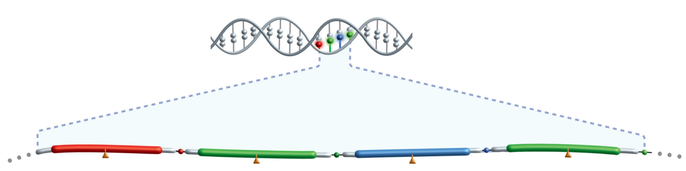
A high-accuracy, high-throughput sequencing technology

Jagadeeswaran Chandrasekar[1], Melud Nabavi[1], Aaron Jacobs[1], Marc Prindle[1], Cynthia Cech[1], Taylor Lehmann[1], Kendall Berg[1], Alan Kimura[1], Alberto Gatto[2], Alec Sautter[1], Alex Lehmann[1], Ali Carroll[1], Amal Chaturvedi[2], Anasha Arryman[1], Bill Maney[2], Brent Banasik[1], Brittany Kesic[1], Charlotte Yang[1], Chen Zhao[2], Daniel Zinder[2], Dieter Heindl[3], Dinendra Abeyawardhane[1], Douglas Lopez[2], Emily Ormbrek[1], Grant Kingsley[1], Hannes Kuchelmeister[3], Hasham Saqib[2], Jason Komadina[2], Geoffrey Barrell[2], Joanne Leadbetter[1], John Mannion[2] , Kevin Smith[1], Lacey McGee[1], Mahdi Golkaram[2], Majid Babazadeh[1], Maryam Shenasa[2], Matthew Lopez[1], McKenna Osentowski[1], Megan Freer[1], Megan LeProwse[1], Megan Rogers-Peckham[1], Michael Kovarik[1], Mitchell Wolfin[1], Robert Busam[1], Ron Cicero[2], Salka Barrett[1], Scott Helgesen[2], Scott Miller[1], Takeshi Harada[2], Thao Ho[1], Thomas Reid[1], Robert McRuer[1], Mark Kokoris[1]

[1]Roche Sequencing Solutions, Inc. Seattle, WA, USA
[2]Roche Sequencing Solutions, Inc. Santa Clara, CA, USA
[3]Roche Diagnostics GmbH, Penzberg, German

Poster #P23.017.C

Introduction to Sequencing By Expansion (SBX)



Single molecule sequencing of nucleic acids is limited by the spatial differentiation of neighboring nucleotides. Sequencing By Expansion (SBX) solves these issues by creating an Xpandomer (Xp) surrogate molecule with >50x separation between high signal-to-noise nucleotide reporters.

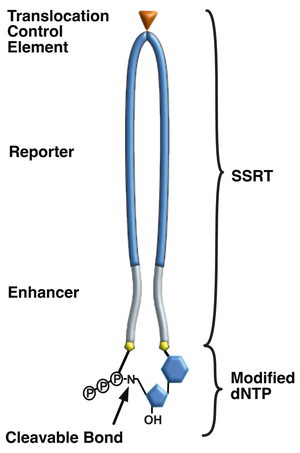


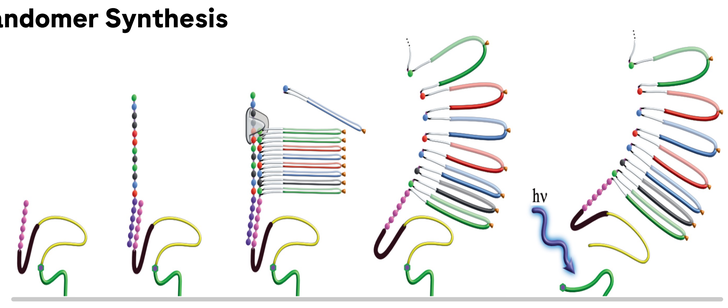
Figure 1. X-NTP Structure

The Expandable Nucleotide Triphosphate (X-NTP) is the building block of the Xpandomer. There are 4 different XNTPs, one for each nucleotide.

The XNTP consists of a modified deoxynucleoside triphosphate (dNTP) that is doubly-conjugated to a symmetrically-synthesized reporter tether (SSRT). Highly designed components of the SSRT allow for high signal to noise and non-stochastic translocation control.

SBX consists of two steps: synthesis and sequencing. First, during synthesis, the DNA template is enzymatically and chemically transcribed into the lengthened Xp molecule that contains high signal-to-noise reporters (**Fig. 2**). Next, the Xp is sequenced by stepwise translocation through a nanopore, (**Fig. 3**) where each step measures a well-defined electrical signal specific to one of four reporters that identifies the DNA base.

Figure 2. Xpandomer Synthesis



(1) A DNA template is hybridized to a specialized, acid-stable primer called an extension oligo (EO), which is anchored to the flow channel. (2) The primed DNA is extended via a highly-engineered DNA polymerase (Xp Synthase) and the four base-specific XNTPs. (3) Acid treatment cleaves each incorporated XNTP, expanding the Xp backbone. (4) The fully formed Xp is released from the solid substrate by photocleavage, and the eluted sample is ready for measurement in the nanopore sequencer.

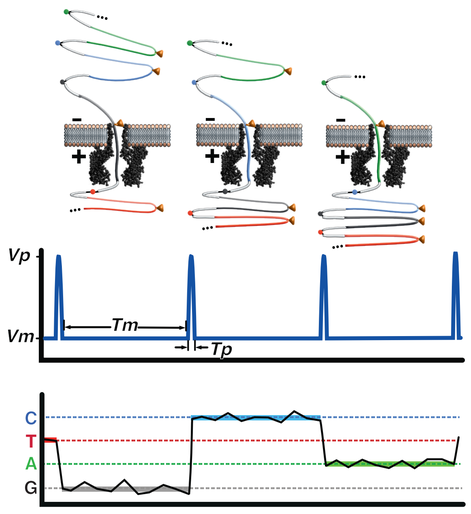


Figure 3. Xpandomer Measurement

The Xp is measured in a nanopore inserted into a lipid bilayer membrane. An applied periodic electrical field is used to drive threading of the Xp through the nanopore.

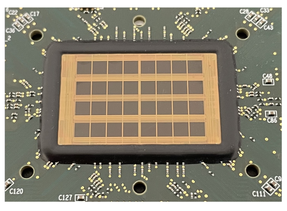
The Xp translocates through the pore until paused by the Translocation Control Element (TCE). The TCE is positioned to hold its adjacent base reporter in the barrel of the nanopore while the electrical signal is measured. A brief, high voltage pulse is used to advance the Xp to the next TCE and subsequent reporter measurement. This cycle continues until the Xp is fully translocated and the nanopore becomes open again to capture another Xp.

Materials & Methods

SBX-D Standard Workflow

50 ng of unsheared genomic DNA from HG001-7 was input into the standard SBX-D library prep workflow. Linear amplification is performed instead of standard PCR, which gives significant advantages, including decreasing error propagation. Following Xp synthesis, the encoded Xpandomers were sequenced as a 7 plex for 1h on an 8M CMOS-based sensor module.

Figure 4. Sensor module



SBX-Fast Amplification-Free Workflow

2 µg of unsheared genomic DNA from HG001-7 was input into a more streamlined SBX-D library prep workflow without amplification. Samples were sequenced as a trio for 1 hr.

SBX Duplex (SBX-D)

Duplex Sequencing By Expansion (SBX-D) is a library prep and sequencing workflow that links both strands of the target DNA in a single sequencing read. This approach uses intramolecular consensus of the complementary reads to identify discordant calls and achieve high accuracy.

Figure 5. SBX-D Standard Workflow

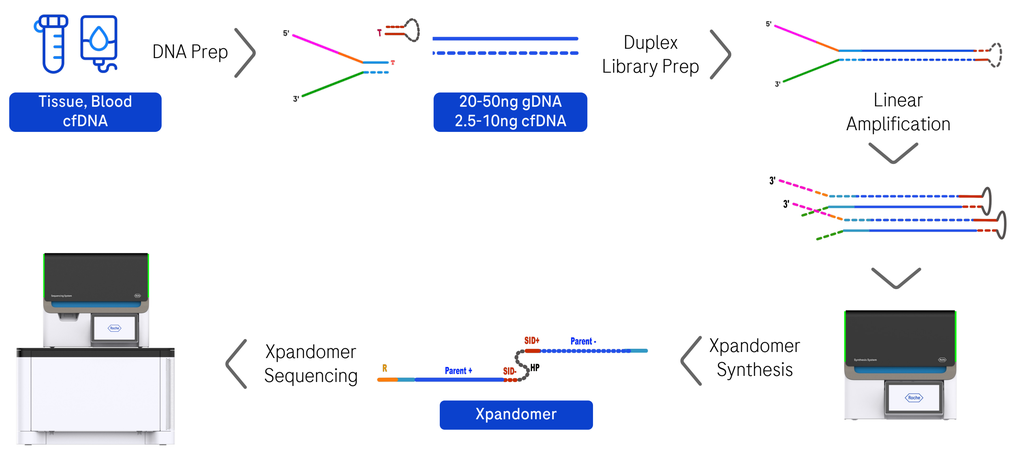
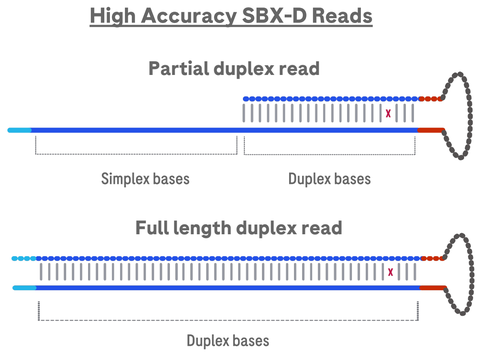


Figure 6. SBX-D Read Categories



Full-length duplex reads contain both the full parent+ and parent- strands. The vast majority of duplex base positions are concordant and marked as high quality. A small percentage of duplex positions will be discordant, and an intramolecular consensus base call will be made, and marked with a low quality score.

Partial duplex also contain simplex stretches of base calls. For these positions in the resulting intramolecular consensus reads, bases are once again called, but are given a medium categorical quality score. On average, the quality of simplex bases is >Q20.

SBX-D base accuracy, coverage and throughput metrics are calculated using ONLY concordant duplex bases.

Results - GIAB

SBX-D and SBX-Fast workflows >99.9% of the genome with at least 10x coverage. F1 scores are calculated on GIAB 4.2.1 high confidence region using all reads from 1 hour of sequencing.

SBX-D Workflow (Linear Amp.)

						GATK + Roche Machine Learning **F1 Score**		*Google DeepVariant* **F1 Score**	
7 plex	Sample	Total Reads	Median Coverage	Mean Insert Read Length	Q score	SNV	Indels	SNV	Indels
GIAB	HG001	0.72	38	244	40.2	99.82%	99.70%	99.87%	99.73%
	HG002	0.76	42	251	41.1	99.78%	99.73%	99.84%	99.75%
	HG003	0.75	41	249	40.6	99.74%	99.63%	99.80%	99.66%
	HG004	0.70	37	246	39.6	99.76%	99.67%	99.83%	99.72%
	HG005	0.70	38	252	40.5	99.76%	99.74%	99.84%	99.77%
	HG006	0.79	44	254	40.6	99.77%	99.77%	99.83%	99.80%
	HG007	0.71	38	248	40.7	99.73%	99.68%	99.80%	99.70%

SBX-Fast Workflow (Amp. free)

						GATK + Roche Machine Learning **F1 Score**		*Google DeepVariant* **F1 Score**	
Trio	Sample	Total Reads	Median Coverage	Mean Insert Read Length	Q score	SNV	Indels	SNV	Indels
CEPH	HG001	1.24	53	190	40.1	99.81%	99.83%	99.86%	99.80%
	NA12891	1.24	52	192	40.4	*Ground truth not available*			
	NA12892	1.23	51	192	40.5				
Ashkenazi	HG002	1.21	56	200	40.1	99.77%	99.82%	99.83%	99.80%
	HG003	1.16	53	199	40.0	99.72%	99.71%	99.78%	99.71%
	HG004	1.12	50	198	39.9	99.75%	99.79%	99.82%	99.80%
Chinese	HG005	1.17	54	203	40.0	99.76%	99.81%	99.83%	99.81%
	HG006	1.10	50	203	40.2	99.75%	99.81%	99.82%	99.83%
	HG007	1.20	53	200	40.2	99.72%	99.74%	99.78%	99.74%

Conclusions

SBX is a transformative sequencing technology that pairs high throughput and high accuracy to improve existing applications and to enable innovation in biological research and molecular diagnostics.

Please contact lead author at: jagdeesh.chandrasekar@roche.com