SEQUENCING NOTHING: EXPLORING FAILURE MODES OF NANOPORE SENSING AND IMPLICATIONS FOR LIFE DETECTION. J. Hachey[1], A. Pontefract[1,2], M.T. Zuber[2], G. Ruvkun[1] and C.E. Carr[1,2]. [1]Deparment of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114. [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 (apontefr@mit.edu).

Introduction: Extant life detection is once again a focus of future planetary missions, and a current challenge is to target biomarkers that can provide unambiguous evidence of life. In this vein, DNA sequencing has previously been suggested as a tool for unambiguous life detection for planetary exploration missions [e.g., 1-2]. Sequencing offers a high level of unambiguity as there are no known abiotic routes to sequence lengths beyond >30 bp [3], and allows for the detection of forward contamination, avoiding false positives. Life detection instruments targeting DNA and/or related polymers are in active development [2,4], however many challenges remain, i.e., use of biological reagents, robust extraction method(s), sequencing non-standard bases or polymers, base-calling algorithms, as well as data analysis and reduction, to cope with data transmission limitations.

The focus of this abstract is on failure modes, specifically the issues encountered when there is no-to-low DNA input into a sequencing device. So called next generation sequencing (NGS) relies upon amplification of this DNA, either during library preparation or the sequencing process, which, especially in the absence of input DNA, produces artifacts (random DNA) and possible machine malfunctions. This is problematic in instances where either very low amounts of DNA must be quantified, or when the absence of any DNA must be accurately determined, i.e. in clean room surveys and in life detection missions.

Here we assess whether single molecule sequencing, involving no amplification whatsoever, generates similar noise signals. To do so, we designed an experiment to clearly elucidate what data are generated when no input DNA is present, using the Oxford Nanopore Technologies (ONT) MinION nanopore sequencer, which is being utilized as part of the SETG (Search for Extra-Terrestrial Genomes) project [4].

Methods: Experiments were conducted using the ONT MinION Mk 1-B with R9 (FL0-MIN104) flowcells. Four cases were explored (**Fig. 1**); in each case the number of 1D and 2D passing and failing reads and bases were recorded, as well as quality (Q)-scores, and mean and median read lengths. Runs 1 and 3, and 2 and 4 were run sequentially on the same respective flow cells for 3-4 hours each. Data was recorded and any passing reads were analyzed using NCBI BLASTN.

Results and Discussion: Runs 1 and 2 established

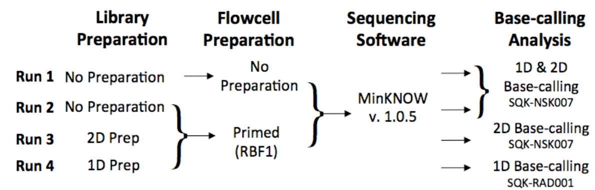

Figure 1. *Workflow for experiment detailing library and flowcell preparation utilized, sequencing software and base-calling analyses. Runs 1-3 used Metrichor v. 1.107 SQK-NSK007, Run 4 used SQK-RAD001.*

Table 1. *Results of four runs showing number of passing and failing bases with corresponding quality (Q) scores.*

	Base-calling Analysis	Measured Bases	Pass Bases	Passing Q-Score (Avg)	Fail Bases	Failing Q-Score (Avg)
Run 1	1D	883,558	487	6.7	883,071	3.9
	2D	596,128	0	-	596,558	4.0
Run 2	1D	1,084,494	9,105	7.3	1,075,389	4.8
	2D	1,246,044	0	-	1,246,044	3.9
Run 3	2D	2,278,106	0	-	2,278,106	2.9
Run 4	1D	1,295,373	192	8.25	1,295,181	3.9

a baseline for the instrument, without the use of library preparation reagents. In both cases no 2D passing reads were detected, and 3 and 2 passing 1D reads were generated respectively. In Run 1, the passing 1D reads contained only one long read, comprised of a series of A's and T's. In Run 2, passing reads were characterized by an over-representation of thymine. Run 3 generated no passing reads, and Run 4 generated 1 passing read for 1D (**Table 1**).

Conclusions: In all cases where a passing read occurred, analysis of the sequence (e.g. nucleotide abundances, or phylogenetic analysis) were clearly detectable as spurious. Thus, with proper filtering, single molecule sequencing produces little noise, and even small numbers of single molecule reads mapping to a organism can provide unambiguous detection of an organism. In previous work, we have shown cases where very low input DNA (i.e. 0.1 ng) is used, and have successfully detected reads that accurately map to the known input [5]. The noise characteristics in all instances suggest that the quality thresholds (Q>6 for 1D, Q>9 for 2D) were appropriately chosen by ONT. New R9.4 chemistry possibly addresses the signals seen in these experiments, but has not yet been tested.

Acknowledgements: NASA award NNX15AF85G.

References: [1] John, K.K. et al. (2016) *LPSC XLVII Abstract #2982.* [2] Carr et al. (2016) *IEEE Aerospace* [3] Smith, K.E. et al. (2016) *Orig. Life. Evol. Bios.* 1-9. [4] Carr et al. (2017) *IEEE Aerospace* (accepted) [5] Mojarro et al. (2017) *LPSC XLVIII Abstract #1585.*