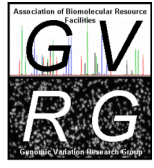


ABRF GVRG 2009 Research Poster

A Review of Heterozygous Bases Detected in *S.Cerevisiae* Using Multiple Next-Gen Sequencing Platforms

C. Lytle, A. Brown, J. Forrester, N. Bivens, B. Sanderson, A. Noll, A. Hutchinson, H. Escobar, M. Detwiler, B. Kingham,

1) Dartmouth College, Hanover, NH, United States, 2) The Partners Center for Personalized Genetic Medicine, Cambridge, MA, United States, 3) University of Missouri, Columbia, MO, United States, 4) Stowers Institute for Medical Research, Kansas City, MO, United States, 5) CGF, NCI, SAIC-Frederick, Gaithersburg, MD, United States, 6) University of Utah, Salt Lake City, UT, United States, 7) Roswell Park Cancer Institute, Buffalo, NY, United States, 8) University of Delaware, Newark, DE, United States.



Abstract

Sequencing methods used by commercial next-generation sequencing platforms produce exponentially more data than Sanger instruments. Yet the quality of these base calls is inherently lower than that of Sanger sequencing. This can make discrimination of genotypic variation difficult, particularly when dealing with heterozygous sites in a diploid organism. The Genomic Variation Research Group (GVRG) 2009 study investigated the current base calling accuracy of two commercial next-generation sequencing platforms and compared these base calls collected to other high- and low-throughput genotyping platforms. A diploid strain of yeast (*Saccharomyces cerevisiae*) sequenced by the Stowers Institute (Li et al., Cell, 2008) was the subject of our study. GVRG collected genotyping data using two next-generation sequencers, the ABI SOLiD and Illumina GAI. We have also included the sequence previously obtained from an Illumina GAI Lower throughput platform, ABI 3730 and 3730XL DNA Analyzers, were used to interrogate regions in or near repetitive sequence that appear to contain a SNP. The Sequenom MassARRAY system, a high-throughput genotyping platform, will be used to assay 900+ previously identified SNPs discovered in the Stowers yeast strain. We will present the comparative SNP data across all platforms and highlight the amount of time, cost of running this sample and different analysis techniques on the different platforms used.

Methods

Yeast DNA extraction:
A single colony of the yeast strain RLY615, a derivative of the S288c wild-type strain, was isolated and inoculated into YPD broth then incubated at 30C for 48 hours. Genomic DNA was prepared from this culture using a phenol-chloroform based extraction technique adapted from Current Protocols in Molecular Biology (2003, Unit 13.11).

Amplicons for Sanger sequencing:
Primers flanking regions of interest were designed and ordered using IDT's online primer design tool. After amplification, the products were digested with exonuclease I and shrimp alkaline phosphatase to prepare them for Sanger sequencing. Amplicons were aliquoted and shipped to GVRG members. Sanger Sequencing was performed using standard ABI chemistry and ABI 3730 and 3730xl sequencers.

Next-gen-sequencing:
Genomic DNA was sent to ICBR, University of Florida for mate-paired library construction and ABI SOLiD sequencing. Illumina single-read library construction was performed by the DNA Core Facility at the University of Missouri. Illumina GAI sequencing was performed by the University of Delaware Sequencing & Genotyping Center and Illumina.

Data Analysis:
SOLID data was analyzed by the ABI SOLiD Bioinformatics specialist using Corona software. This data was also filtered for coverage depths between 3 and 255 bases. Maq software was used in conjunction with an in-house filter developed by Stowers Institute Bioinformatics group. All next-gen sequencing data was analyzed using this software.

Results

When comparing the heterozygous calls between both systems it is striking at the amount of calls that that one system makes that the other did not. There are multiple factors that may be contributing to this. These factors likely include the SOLiD mate-paired versus the Illumina single-read library. The biggest factor appears to be differences between the two analysis programs. Corona and Maq, and how they process input data. Running the SOLiD data through Maq brings the total number of heterozygous calls more in-line with those from the Illumina systems, but does not significantly increase the heterozygous calls agreed upon by the two systems. The GVRG will attempt to verify calls made by these systems using a combination of Sequenom assays and allelic discrimination assays for many of the sites.

>SOLID data analyzed with Corona identified 1271 SNPs, of these 253 are heterozygous

>SOLID data analyzed with Maq identified 906 SNPs, of these 97 are heterozygous

>Illumina GAI analyzed with Maq identified 1115 SNPs, of these 166 are heterozygous

>Illumina GAI analyzed with Maq identified 1035 SNPs, of these 73 are heterozygous

>There are 809 SNPs in agreement among the systems, of these 23 are heterozygous (this does not include SOLiD/Maq data set)

>A total of 441 individual Heterozygous were identified when combining all four data analysis sets (listed here)

We would like to thank

>The Stowers Institute for providing the Yeast cell line.
>Chares Cochran of Applied Biosystems for organizing the SOLiD assay and analysis.
>Applied Biosystems and providing materials for the SOLiD assay.
>Jingwei Ni of Applied Biosystems for the SOLiD data analysis.
>Bill Farmerie and the ICBR lab for running the SOLiD assays.
>IDT for providing over 2000 oligos at very discounted prices.

Chrom	Position	Ref	SOLID Cor	SOLID Maq	Illum GA2	Illum 1G	Seque	Sanger	Chrom	Position	Ref	SOLID Cor	SOLID Maq	Illum GA2	Illum 1G	Seque	Sanger	Chrom	Position	Ref	SOLID Cor	SOLID Maq	Illum GA2	Illum 1G	Seque	Sanger			
I	968	C	Y	75	51	Y	136		II	75929	C	M	180					XIII	317129	T	Y	13							
I	1025	C	Y	235	70	Y	72		II	75989	G	R	46					IX	333323	T	Y	13							
I	2485	A	W	70	70	W	60		IX	843214	C							IX	289322	T			W	70					
I	2485	A	W	70	70	W	60		IX	866910	T							XIII	391110	A			W	70					
I	3835	A	W	60	104				IX	978177	A							XIII	391135	A			R	65					
I	12739	T	W	60					IX	986047	A	W	6					XIII	391144	T			K	241					
I	12765	T	Y	100					IX	992702	A	R	127					XIII	391163	A			R	78					
I	13299	C	Y	256					IX	1003629	A							XIII	391201	A			R	205					
I	25419	T	Y	51	146				IX	1101681	T	Y	115					IX	435031	G			K	33	238				
I	25463	G	K	54					IX	1307938	G		R	176				IX	439848	A			R	151					
I	25468	G	K	54					IX	1379034	T							X	441	G			S	66					
I	25489	G	R	49					IX	1456193	T							X	493	Y			Y	105					
I	25516	A	R	9					IX	1502734	A							X	591	C			Y	48					
I	25612	G	R	24					IX	1525390	G							X	673	C			Y	43					
I	25615	Y	Y	27					IX	1529416	G							X	701	T			Y	53					
I	25621	G	R	40					IX	1529526	T							X	724	S			Y	20					
I	25624	G	R	48					IX	1525432	T	Y	16					X	60986	G			K	16					
I	25711	A	W	9					IX	1645	T							X	7544	C									
I	25714	G	R	5					IX	100605	T							X	9780	Y			Y	39	118				
I	25852	C	Y	12					IX	226377	T							X	247047	C			S	20					
I	26087	C	Y	8					IX	303033	G							X	27864	A			K	27					
I	26222	C	Y	30					IX	400098	A							X	28900	G									
I	26242	A	R	30					IX	570408	G							X	1883	A			M	31					
I	26254	A	R	11					IX	8153	C	Y	34					X	399654	A			K	36					
I	26389	A	R	8					IX	12416	T	W	74					X	406539	T			K	36					
I	26392	A	R	12					IX	97903	A	M	16					X	477468	R			R	24					
I	26800	G	R	31					IX	106271	C							X	477667	S			S	64					
I	27100	T	Y	8					IX	18139	A	M	51					X	477670	A			M	73					
I	27127	Y	Y	18					IX	213371	A							X	477700	G			R	12					
I	45635	C	Y	18	24				IX	270146	T	K	167					X	478556	A			M	57					
I	98350	C	Y	50	24				IX	72199	G							X	42914	A			R	30					
I	98351	G	Y	41					IX	72110	A	W	59					X	713209	T			W	187					
I	126881	C	R	6					IX	122568	G	R	227					X	714892	G			R	117	255				
I	181390	C	R	6					IX	125959	G	S	28					X	714919	G			R	59					
I	181402	G	R	15					IX	239625	T							X	714921	C			Y	61					
I	171882	G	M	16					IX	286725	G	K	16					X	71504	G			R	56					
I	188875	G	S	51	142				IX	384847	G	S	53					X	715054	A			Y	69					
I	188933	A	R	255	52				IX	404477	A	R	19					X	715089	G			R	22					
I	18911	C	Y	7					IX	415088	A	S	41					X	71524	R			X	427150					
I	204896	C	Y	7					IX	413116	A							X	71525	G			K	9					
I	204874	C	Y	13					IX	413763	A							X	71526	A			R	5					
I	204889	G	R	11					IX	419056	G	R	51					X	73809	K			K	78	220	K	67	A	21
I	205141	G	R	8					IX	509490	A							X	742895	A			W	8					
I	205141	G	R	8					IX	530023	A	M	97					X	745334	A			R	30					
I	205629	G	R	12					IX	530308	A							X	745334	A			R	30					
I	206369	C	Y	75	255				IX	622718	T							X	745334	A			R	30					
I	206147	T	W	10					IX	655686	T							X	75131	C			S	43					
I	206360	T	Y	32					IX	678074	C							X	7132	G			S	40					
I	206366	C	Y	76	110				IX	781704	T	K	34					X	72941	G			K	63					
I	206612	T	Y	64	138				IX	817482	A	R	22					X	147890	A			K	32					
I	206613	G	C	64	138				IX	817495	G	S	48					X	239023	T			K	37					
I	206624	C	Y	3					IX	866215	A							X	317770	A			R	12					
I	210454	C	R	38					IX	919964	G	M	74	83				X	454840	T			W	6					
I	210471	A	R	38					IX	960516	T							X	461922	A			M	18					
II	5490	A	R	63																									