Abstract # 231495

Seventeen percent of NGS 50 gene panel variants are not expressed in RNAseq

CONTRIBUTING RESEARCHERS

Razelle Kurzrock[1], Rahul Parulkar[2], Timothy Joseph Yeatman[3], Wafik S. El-Deiry[4], Timothy J. Pluard[5], Chad Garner[6], Sandeep K. Reddy[7]

[1]Moores Cancer Center, La Jolla, CA; [2]NantOmics LLC., Santa Cruz, CA; [3]Gibbs Cancer Center and Research Institute, Spartanburg, SC; [4]Fox Chase Cancer Center, Philadelphia, PA; [5]St. Luke's Cancer Institute, Kansas City, MO; [6]NantOmics LLC., Culver City, CA; [7]NantHealth LLC., Culver City, CA

BACKGROUND

- Tumor-only sequencing analysis to identify somatic variants increases the risk of mistakenly identifying germline mutations as somatically-derived cancer mutations.
- Simultaneous bioinformatics analysis of both the normal germline and tumor genome along with RNA analysis is necessary for accurate identification of molecular targets for cancer therapy.
- Standard NGS panels evaluate DNA only. RNAseq has shown that molecular targets identified by NGS panels are not universally expressed.
- The objective of this study was to compare the accuracy and precision of tumor somatic calling with a 50 gene commonly used hotspot panel, analyzing tumor tissue alone versus analyzing tumor DNA simultaneously with normal germline DNA and tumor RNA.
- Furthermore, we hypothesized that heterogeneous epigenomic factors may lead to **low or absent RNA expression**. We sought to determine the frequency of **non-expressed variants** that would be tested by a standard NGS panel.

METHODS

- This study included 1879 cancer patients with 42 cancer types with either whole genome sequencing or whole exome sequencing of both tumor and normal genomes.
- True positive (true somatic variants) and false positive (true germline variants estimated to be somatic variants) rates were measured for **missense** and **nonsense** single nucleotide variants (SNVs) in a 50 gene panel.
- A in-silico 50 gene panel (Ampliseq HotSpot V2) was constructed as a reference comparison: ABL1, EGFR, GNAS, KRAS, PTPN11, AKT1, ERBB2, GNAQ, MET, RB1, ALK, ERBB4, HNF1A, MLH1, RET, APC, EZH2, HRAS, MPL, SMAD4, ATM, FBXW7, IDH1, NOTCH1, SMARCB1, BRAF, FGFR1, JAK2, NPM1, SMO, CDH1, FGFR2, JAK3, NRAS, SRC, CDKN2A, FGFR3, IDH2, PDGFRA, STK11, CSF1R, FLT3, KDR, PIK3CA, TP53, CTNNB1, GNA11, KIT, PTEN, VHL.
- RNA sequencing was available for **1134/1879 (60%)** patients.
- Sequence alignment and SNV variant calling was performed using well-established and published bioinformatics methods (References).
- Post alignment statistics for RNA were confirmed to contain at least 10x coverage on average of 300 key genes known to have a role in cancer.

METHODS REFERENCES

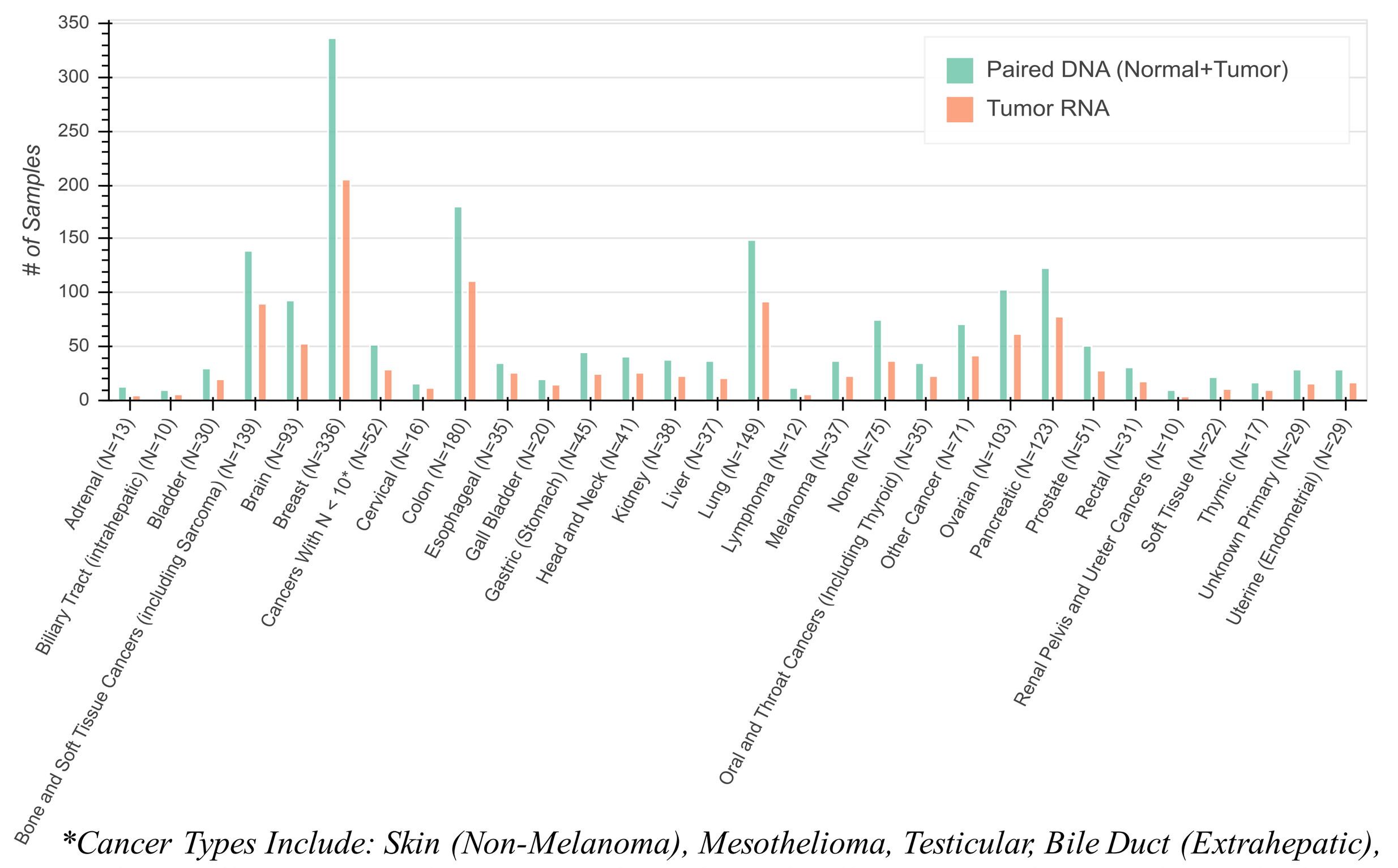
Sanborn JZ, Chung J, Purdom E, Wang NJ, Kakavand H, Wilmott JS, Butler T, Thompson JF, Mann GJ, Haydu LE, et al.: **Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination**. *Proc Natl Acad Sci U S A* 2015, **112**:10995-11000.
Schulman JM, Oh DH, Sanborn JZ, Pincus L, McCalmont TH, Cho RJ: **Multiple Hereditary Infundibulocystic Basal Cell Carcinoma Syndrome Associated With a Germline SUFU Mutation**. *JAMA Dermatol* 2016, **152**:323-327.
Shain AH, Garrido M, Botton T, Talevich E, Yeh I, Sanborn JZ, Chung J, Wang NJ, Kakavand H, Mann GJ, et al.: **Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE promoter mutations and diverse activating mutations in the MAPK pathway**. *Nat Genet* 2015, **47**:1194-1199.
Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**:1754-1760.
Faust GG, Hall IM: **SAMBLASTER: fast duplicate marking and structural variant read extraction**. *Bioinformatics* 2014, **30**:2503-2505.
DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al.: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nat Genet* 2011, **43**:491-498.
Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biol* 2009, **10**:R25.
Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *BMC Bioinformatics* 2011, **12**:323.
Monkol Lek, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, et al. **Analysis of protein-coding genetic variation in 60,706 humans** *Nature* **volume 536**, pages 285–291 (18 August 2016) doi:10.1038/nature19057
Teer JK, Zhang Y, Chen L, et al: **Evaluating somatic tumor mutation detection without matched normal samples**. *Hum Genomics* 2017, **11**:22.
Mandelker D, Zhang L, Kemel Y, et al: **Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline based germline testing.** JAMA 2017, 318:82
Jones S, Anagnostou V, Lytle K, et al.: **Personalized genomic analyses for cancer mutation discovery and interpretation**. *Sci Transl Med* 2015, 7:283

RESULTS

Table 1. Demographics overview for cohort (N=1879).

Cancer Type	# Patients	# Male	# Female	Min. Age	Max Age	Median Age
Breast	336	2	327	20	86	56
Colon	180	83	93	17	87	58
Lung	149	67	78	9	90	65
Bone and Soft Tissue Cancers (including Sarcoma)	139	72	62	0	82	49
Pancreatic	123	69	48	3	87	63
Ovarian	103	0	96	25	86	58
Brain	93	52	37	0	79	49
Cancer Type Unknown	75	38	29	6	91	59
Other Cancer	71	39	31	1	83	62
Cancers With N < 10*	52	29	20	0	87	65.5
Prostate	51	48	0	40	83	65
Gastric (Stomach)	45	26	19	15	85	61
Head and Neck	41	31	8	19	86	64
Kidney	38	25	11	0	72	62
Liver	37	25	11	9	77	63
Melanoma	37	24	12	29	87	64
Oral and Throat Cancers (Including Thyroid)	35	21	13	42	83	63
Esophageal	35	24	10	46	86	64
Rectal	31	21	10	28	80	57
Bladder	30	17	12	49	92	72
Unknown Primary	29	11	18	29	83	57
Uterine (Endometrial)	29	0	28	34	89	66
Soft Tissue	22	15	7	2	80	18
Gall Bladder	20	7	13	39	87	65.5
Thymic	17	9	8	24	73	59
Cervical	16	0	16	27	75	49
Adrenal	13	8	4	1	74	48
Lymphoma	12	8	3	18	81	66
Renal Pelvis and Ureter Cancers	10	5	5	8	71	42
Biliary Tract (intrahepatic)	10	5	4	46	78	61

Fig 1. Analytes sequenced per cancer type.



**Cancer Types Include: Skin (Non-Melanoma), Mesothelioma, Testicular, Bile Duct (Extrahepatic), Anal, Ampulla of Vater, Leukemia, Vaginal, Myeloma, Small Intestine, Vulvar, Penile, Urethral*

- Sequencing the tumor genome identifies all of the SNVs of inherited germline origin and tumor somatic origin, and the large majority are of germline origin. (Fig. 2)
- Population allele frequencies and other parameters can be used to filter SNV data and estimate somatic versus germline origin, although not accurately enough for clinical use, as shown recently by others. (Fig. 3)
- All but 1 true germline mutation is identified in < 1% of samples. (Fig. 3)

Fig 2. 92% of SNVs identified from sequencing tumor genomes alone were of germline origin and potential false positives rather than true somatic variants.

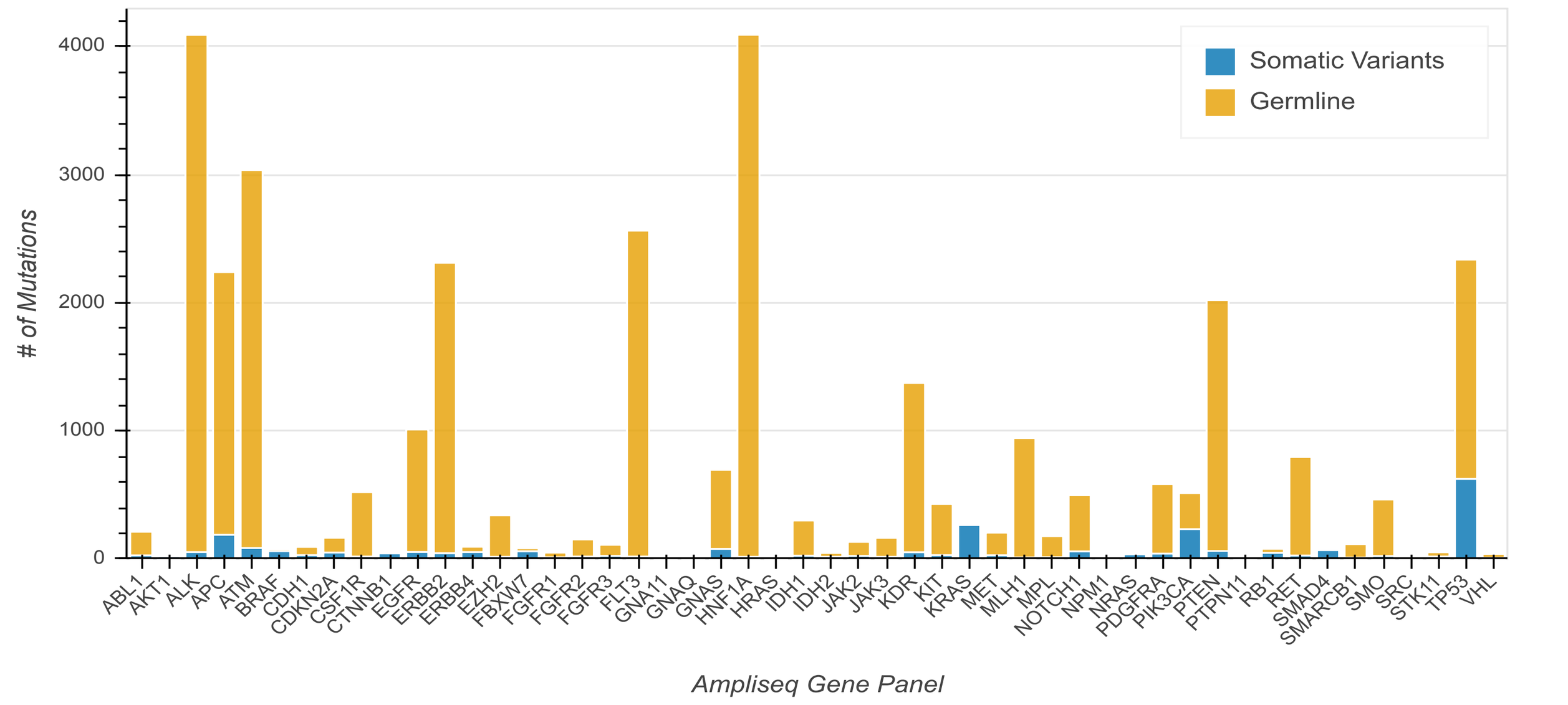


Fig 3. Filtering all SNVs using gnomAD with reported population allele frequencies >= 0.001 OR whether the variant existed at all still resulted in a false positive rate of 34%.

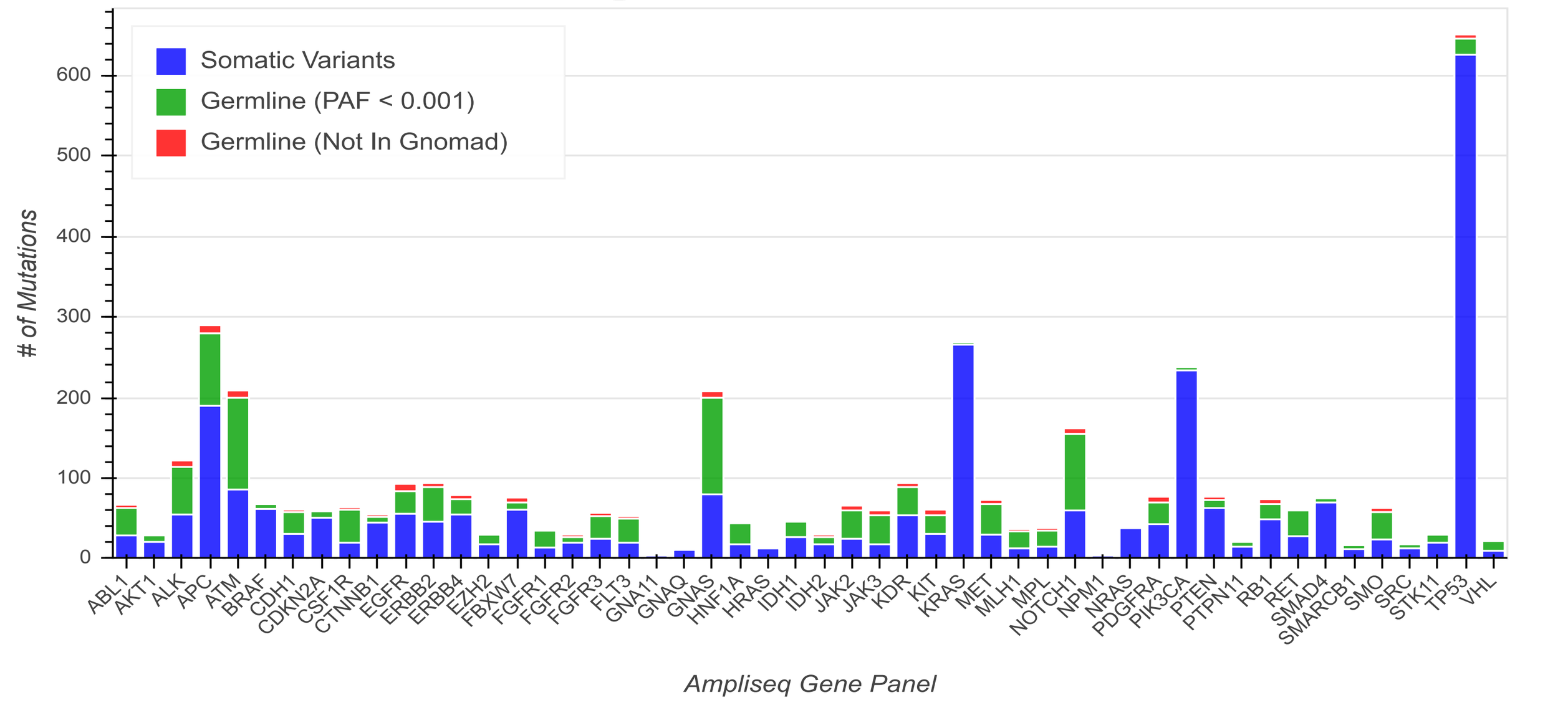


Fig 4. RNA analysis showed that 15% of somatic SNVs (missense/nonsense) and 17% of all somatic SNVs (missense/nonsense/synonymous) were not expressed. 23% of patients had at least one somatic variant (misense/nonsense) that was not expressed.

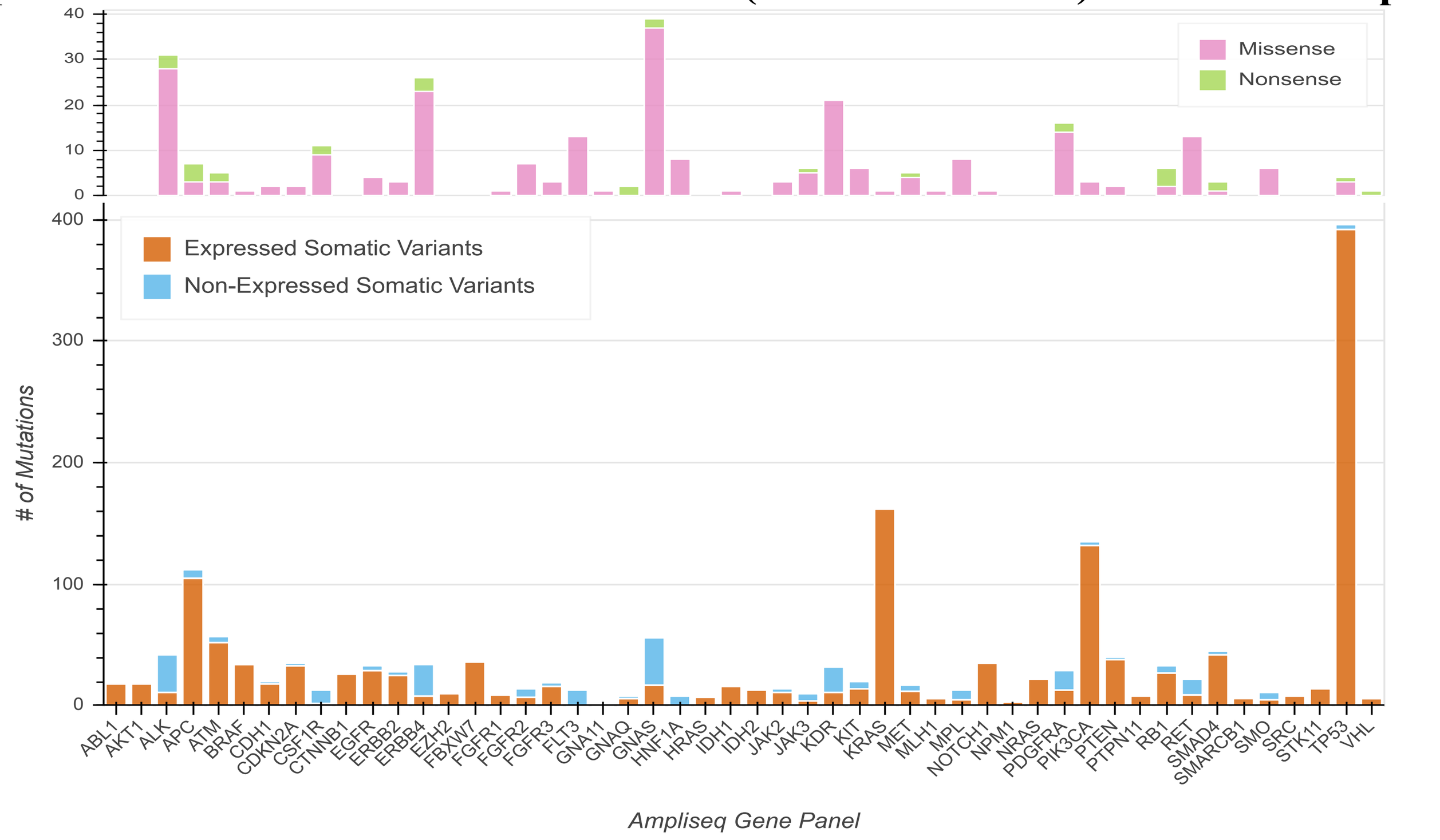


Table 2. Somatic SNVs in gene panel per cancer type, sorted by their respective expression percentages.

Cancer Type	# Mutations in Panel	Unique Mutations in Panel	Expressed Mutations in Panel	% Expressed	Average Non-Synonymous Load
Melanoma	76	70	44	58%	546.54
Skin (Non-Melanoma)	67	67	45	67%	2660.33
Gall Bladder	13	11	9	69%	77.45
Lymphoma	4	4	3	75%	134.92
Kidney	12	12	9	75%	344.61
Leukemia	4	4	3	75%	115.75
Soft Tissue	8	6	6	75%	116.45
Head and Neck	28	28	22	79%	120.71
Cancer Type Unknown	149	145	117	79%	437.24
Biliary Tract (intrahepatic)	5	5	4	80%	71.9
Vaginal	5	5	4	80%	498.33
Thymic	5	5	4	80%	155.82
Bile Duct (extrahepatic)	5	5	4	80%	59.6
Esophageal	45	40	37	82%	169.6
Mesothelioma	6	6	5	83%	52.25
Prostate	19	19	16	84%	95.18
Bone and Soft Tissue Cancers (including Sarcoma)	56	54	47	84%	123.81
Liver	19	19	16	84%	160.22
Other Cancer	47	45	40	85%	105.82
Bladder	48	42	41	85%	235.7
Unknown Primary	13	13	11	85%	160
Lung	168	147	143	85%	261.6
Anal	7	7	6	86%	250.4
Colon	332	230	289	87%	270.22
Oral and Throat Cancers (Including Thyroid)	40	38	35	88%	146.31
Uterine (Endometrial)	28	26	25	89%	180.76
Pancreatic	118	63	105	89%	70.54
Gastric (Stomach)	29	28	26	90%	136.58
Breast	225	154	202	90%	124.83
Ovarian	52	46	48	92%	100.72
Brain	59	46	56	95%	90.61
Rectal	43	39	41	95%	241.29
Cervical	26	24	25	96%	229.38
Adrenal	2	2	2	100%	85.62
Small Intestine	2	2	2	100%	241
Testicular	1	1	1	100%	393.17
Renal Pelvis and Ureter Cancers	2	2	2	100%	125.8
Myeloma	1	1	1	100%	146
Penile	4	4	4	100%	476
Ampulla of Vater	2	2	2	100%	108

Across 39 tumor types, the range of expression was 58% (Melanoma)-100% (Penile). Urethral/Vulvar cancer types are absent in this table, patients in those cancer types had no mutations in the gene set.
Non-Synonymous Load refers to the raw counts of non-synonymous mutations.

CONCLUSION

- All patients have at least 1 germline SNV (30955 total) in the panel.
- **1227/1879 (65%)** of patients have at least 1 somatic SNV (308721 total) in the panel.
- **741/1135 (65%)** of paired DNA/RNA patients have at least 1 somatic SNV (198844 total) in the panel. This results in **1775 unique SNVs** amongst paired DNA/RNA patients.
- **1502/1775 (84.6%)** were expressed in the RNAseq.
- Upon adding **synonymous mutations, 1673/2031 (82.3%)** were expressed in the RNAseq.
  - **273/351 (76.3%)** of the non-expressed SNVs are **missense/nonsense**.

- Simultaneous sequencing and bioinformatics analysis of DNA, both the normal germline genome and tumor genome, is necessary for accurate identification of molecular targets.
- Analysis of tumor genome alone results in false-positive somatic variant calls.
- Higher precision is achieved with simultaneous tumor-normal DNA and tumor RNA sequencing analysis.
- The lack of RNA expression may contribute to less than expected clinical benefit with molecularly targeted therapies.
- Since the distribution is non-uniform, identification of these genes can yield improved testing algorithms and treatment strategies.

CONTACT

Corresponding Author: rahul.parulkar@nantomics.com

ASCO ANNUAL MEETING, CHICAGO, IL, JUNE 1-5 2018