



Analytical approaches for the BOAC SNP panel association with progression free survival in myeloma

Brian Van Ness¹ Christine Ramos¹, Majda Haznadar¹, John Crowley², Bart Barlogie³, Brian Durie⁴, Antje Hoerring², Jeff Haessler², Vipin Kumar¹, Michael Steinbach¹, David Johnson⁵, Gareth Morgan⁵.

¹University of Minnesota, ²Cancer Research and Biostatistics, ³University of Arkansas Medical Sciences, ⁴Aptium Oncology, ⁵The Royal Cancer Hospital

ABSTRACT

The Bank On A Cure (BOAC) has established DNA banks from multiple cooperative and institutional clinical trials, and platforms for examining the association of genetic variations (SNPs) with disease risk and outcomes in myeloma. We have previously described the development and content of a novel custom SNP panel that contains 3,404 SNPs in 983 genes, representing cellular functions and pathways that may influence disease response, toxicities, complications, and survival. Although survival certainly varies according to tumor heterogeneity (ie. chromosomal abnormalities, gene expression variations) germline variations that influence the microenvironment, drug distribution, drug transport and metabolism, may also have an association with survival outcomes. To explore SNP associations with progression free survival (PFS) we compared the BOAC SNP profiles of short term PFS (less than 1 year, n=70) versus long term PFS (greater than 3 years, n=73) in two phase III clinical trials (ECOG E9487 and SWOG S9321). A variety of analytical approaches was undertaken including univariate rank ordering, recursive partitioning, and support vector machine learning tools (SVM). Each of these approaches has advantages and limitations in dealing with type I false positive errors as well as sensitivity and specificity. We included subset validation approaches and randomization of classes to address how robust and predictive different approaches were. From our analysis we conclude germline genomic variations do have an impact on progression free survival, with a subset of SNPs from the panel reaching 76% predictive association and odds ratios of survival of 9.6 (CI 4.5, 20.5), p<0.001, using SVM analysis. Based on univariate approaches, we find the most significant variations associated with survival differences were genes that could be functionally categorized as pharmacologic. The presentation will focus on the analytical approaches, and refinements necessary to assure predictive value compared to random associations. Notwithstanding the clear importance of tumor cell variations in genetic deregulation, we conclude that various functions within the bone marrow and drug response likely interplay as a complex influence on disease progression, response, and survival.

CHIP DESIGN

• Bank On a Cure (BOAC) Mission: to create a DNA bank and develop genetic correlates with myeloma risk, progression, response and toxicities associated with therapies.

Fig 1. An approach taken in designing the custom SNP chip



Table1. Functional categories on the SNP chip panel

Functional Category	#Genes	#SNPs
ADME/DMET	130	455
Cancer	406	1538
Carbohydrate Metabolism	69	384
Cell Cycle	230	867
Cell Death	433	1662
Cell Signaling	93	352
Cell-To-Cell Signaling and Interaction	248	880
Cellular Growth and Proliferation	420	1451
Cellular Movement	227	923
DNA Replication, Recombination, and Repair	204	854
Drug Metabolism	20	114
Gene Expression	240	951
Hematological Disease	223	876
Immune Response	247	985
Lipid Metabolism	146	664
Molecular Transport	170	708
Nucleic Acid Metabolism	30	161
Skeletal and Muscular Disorders	64	289
Skeletal and Muscular System Development and Function	77	273
Signaling Kinase, Phosphatase, Transferase	198	885
Inflammation & Immunity	196	813

3404 SNPs in total of 983 genes

METHODS

- Two phase III clinical trials (similar therapies): ECOG E9487 / SWOG/ECOG S9321
- SNP associations with progression free survival (PFS): two arms long >3 years (n=73); short <1 year (n=70)

STATISTICAL APPROACHES

- **Univariate Analysis:** single SNP tested against the phenotype for correlation and ranked.
 - Cannot identify interacting SNPs
- **Multivariate Analysis:** groups of SNPs of size two or more are tested for possible association with the phenotype. Example: MDR
 - Often infeasible in practice due to combinatorial explosion.
- **Classification methods:** use the entire set of SNPs as features to build a model for differentiating cases from controls.

Leave One Out Cross validation

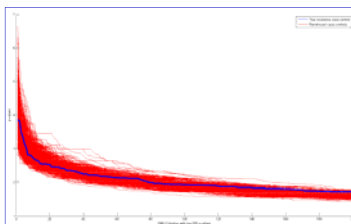
- Each time a sample is left out and a model is built on the remaining data
- The built model is evaluated on the left out sample
- Best suited when the number of samples are small

Permutation Testing

- Labels are shuffled for each evaluation
- This process is repeated 10,000 times
- The p-value is determined based on the number of times the permutation accuracy has crossed the accuracy obtained over actual labels

RESULTS

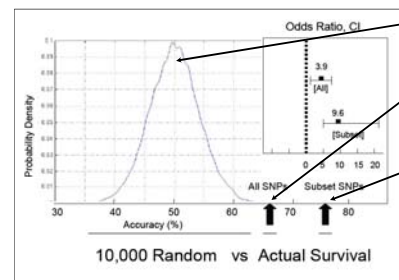
Univariate Permutation Test – P-Value



- Individual SNP associations with true phenotype are not distinguishable from random permutation of phenotype
- A combination of SNPs may be more predictive than individual SNPs

Performance for Selected Categories of SNPs

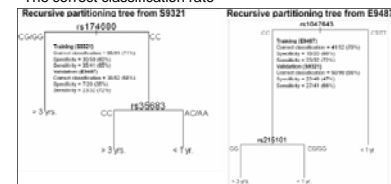
Non-synonymous	Introns	Synonymous	Admixture	UTR	Other intergenic	Accuracy (%)
✓	✓	✓	✓	✓	✓	66.43
✓	✓	✓	✓	✓	✓	58.74
✓	✓	✓	✓	✓	✓	51.74
✓	✓	✓	✓	✓	✓	72.72
✓	✓	✓	✓	✓	✓	71.33
✓	✓	✓	✓	✓	✓	54.54
✓	✓	✓	✓	✓	✓	69.99



- 10,000 random case-control permutations center at 50% accuracy; never above 63%
- Actual survival comparison: all SNPs reached accuracy of 67% in predicting long versus short survival.
- Subgroup of Non-syn + promolign (regulatory) SNPs reached 76% accuracy
- P < .0001

Recursive partitioning

- Recursive partitioning performed on top 50 rank ordered SNPs for each trial separately
- Identify combinations of SNPs that best distinguish PFS groups
- Each genotype is evaluated on its ability to make a correct prediction, creating a decision node
- A pruned decision tree is created in which the minimum number of the strongest nodes creates a group prediction
- The correct classification rate



- Rs174680 Catechol-O-methyltransferase
- Rs35683 Ghrelin
- Rs1047643 Farnesyl transferase 1
- Rs215101 ATP binding cassette C (ABCC)

CONCLUSIONS

- Univariate analysis shows associations for individual SNPs within random permutations
- Classification methods that look at all SNPs or subgroups show association with survival with accuracy significantly above random
- Recursive partitioning shows some weakly associated individual SNPs
- Interactions of SNPs suggest multiple germline gene variations may contribute to survival
- Patient outcomes are likely affected by tumor variations and germline variations

ACKNOWLEDGMENTS

- Van Ness, B. et al. Genomic variation in myeloma: design, content, and initial application of the Bank On A Cure SNP Panel to detect associations with progression-free survival. BMC Med 6, 1-14 (2008).
- Support from the International Myeloma Foundation and NCI P01 51112

No conflicts of interest to report for any author.

