Oral Oncology 47 (2011) 951–955

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Personalized medicine for mucositis: Bayesian networks identify unique gene clusters which predict the response to gamma-D-glutamyl-L-tryptophan (SCV-07) for the attenuation of chemoradiation-induced oral mucositis

Gil Alterovitz [a], Cynthia Tuthill [b], Israel Rios [b], Katharina Modelska [b], Stephen Sonis [c,*]

[a] Children's Hospital Informatics Program at the Harvard/MIT Health Sciences and Technology Division, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Partners HealthCare Center for Personalized Genetic Medicine, United States
[b] SciClone Pharmaceuticals, Foster City, CA, United States
[c] Biomodels, LLC, Watertown, MA, United States

ARTICLE INFO

Article history:
Received 14 March 2011
Received in revised form 14 June 2011
Accepted 9 July 2011
Available online 6 August 2011

Keywords:
Mucositis
Personalized medicine
Genomics
Gamma-D-glutamyl-L-tryptophan
Radiation

SUMMARY

Gamma-D-glutamyl-L-tryptophan (SCV-07) demonstrated an overall efficacy signal in ameliorating oral mucositis (OM) in a clinical trial of head and neck cancer patients. However, not all SCV-07-treated subjects responded positively. Here we determined if specific gene clusters could discriminate between subjects who responded to SCV-07 and those who did not.

Microarrays were done using peripheral blood RNA obtained at screening and on the last day of radiation from 28 subjects enrolled in the SCV-07 trial. An analytical technique was applied that relied on learned Bayesian networks to identify gene clusters which discriminated between individuals who received SCV-07 and those who received placebo, and which differentiated subjects for whom SCV-07 was an effective OM intervention from those for whom it was not.

We identified 107 genes that discriminated SCV-07 responders from non-responders using four models and applied Akaike Information Criteria (AIC) and Bayes Factor (BF) analysis to evaluate predictive accuracy. AIC were superior to BF: the accuracy of predicting placebo vs. treatment was 78% using BF, but 91% using the AIC score. Our ability to differentiate responders from non-responders using the AIC score was dramatic and ranged from 93% to 100% depending on the dataset that was evaluated. Predictive Bayesian networks were identified and functional cluster analyses were performed. A specific 10 gene cluster was a critical contributor to the predictability of the dataset.

Our results demonstrate proof of concept in which the application of a genomics-based analytical paradigm was capable of discriminating responders and non-responders for an OM intervention.

Introduction

Oral mucositis (OM) is a significant toxicity of chemoradiation used to treat head and neck cancer (HNC).[1] γ-D-glutamyl-L-tryptophan (SCV-07), an immunomodulating peptide, reduced the severity and duration of chemoradiation-induced OM in animals.[2] Subsequently, a phase 2a clinical trial suggested that SCV-07 favorably altered the course of OM in HNC patients receiving chemoradiation.[3,4] While the aggregate data was favorable for SCV-07, the response of the treated population was not uniform; only some subjects responded to the drug.

Genomic analysis has shown that genetics can play a key role in drug responsiveness.[7,8] Here, we tested the hypothesis that differences in gene expression could be used to discriminate SCV-07 OM responders from non-responders. Learned Bayesian networks[5,6] identified unique gene clusters which discriminated between individuals who received SCV-07 and those who received placebo, and which differentiated subjects for whom SCV-07 was effective as an OM intervention from those for whom it was not.

This analysis paradigm differs from conventional genomic analyses: it did not mandate a threshold gene expression level change for inclusion, it evaluated simultaneous expression profiles (clusters or networks) of genes rather than individual genes, and the networks (clusters) that evolved were driven by the patterns of gene expression of the study groups, not by any hypothesized expectation. Our method allowed us to evaluate gene expression differences in a way which was not prejudiced by any preconceptions of an expected result. However, we were still able to evaluate the fit of the discovered networks within known ontological pathways. Our results demonstrate the potential of using differences in

* Corresponding author. Address: Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, United States.
E-mail addresses: Gil@chip.org (G. Alterovitz), ssonis@partners.org (S. Sonis).

1368-8375/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.oraloncology.2011.07.006

Table 1
Demographic comparison of subjects with group assignments who provided paired specimens.

	Age (years)	Gender (% ♀)	Ethnicity (% Hispanic)	Race (% Caucasian)	BMI
Placebo	59	12.5 (1/8)	12.5 (1/8)	100	32.3
Non-responders	54	33 (3/8)	17 (1/8)	100	30.5
Responders	54	17 (1/6)	0 (0/8)	100	26

Table 2
Classification of observations used in analysis. Fifty-one observations (samples) were used in the analysis. Of these, 46 were paired samples (samples from the same subject obtained at the screening visit and on the last day of radiation). The remainder originated from subjects for which only one sample from one time was available. While these could not be used to compare specific changes over time, they were of value to contributing to our knowledge of either the screening or LDRT time points. Non-paired samples consisted of one for only screening and four for the LDRT.

	High dose SCV-07	Low dose SCV-07	Placebo	Unassigned	Total
Placebo			16		16
Non-responder	6	12			18
Responder	12	4		1	17
Total	18	16	16	1	51

gene cluster expression as the basis for personalization of OM interventions.

Materials and methods

Subjects

RNA was obtained from HNC subjects (age range 42–84 years) being treated with chemoradiation and enrolled in a clinical trial of SCV-07 (NCT00756951). SCV-07 or placebo was administered subcutaneously on days of radiation.[3,4] The study was approved by the Human Studies' Review Board at each of 18 study sites. OM was assessed using WHO criteria[9] twice weekly throughout the radiation course.

Subjects were categorized to three groups: placebo-treated, SCV-07-treated responders, and SCV-07-treated non-responders. Responders were defined as those who were free of ulcerative OM (WHO grade <2) from their first dose of radiation through a cumulative dose of 40 Gy; non-responders developed ulcerative OM (WHO grade $\geqslant 2$) prior to a cumulative dose of 40 Gy. No significant differences in demographics, tumor stage, or prior surgery were noted among placebo or responder/non-responder cohorts (Table 1).

Not all clinical trial ($n = 57$) enrollees consented to participate in the gene analysis component of the study. Paired RNA specimens, with samples from both the screening visit (prior to chemoradiation or study drug treatment), and last day of radiation (LDRT) from the same individual were available for 23 subjects enrolled in the study (Table 2).

Samples and processing

RNA was isolated from whole blood samples in standard fashion using PAX-gene technology (PreAnalytix, Hombrechtikon, Switzerland). Typical yields of RNA were between 5 and 25 μg. After RNA characterization and quality assurance, microarrays were performed by Expression Analysis, Durham, NC using Illumina Sentrix BeadChip® HumanHT-12 Expression BeadChips containing approximately 48,000 probes (Illumina, Inc., San Diego, CA).

Data analysis

Data analysis was performed in a five stage sequence which included data preparation, selection of predictive genes (attribute selection), Bayesian network probabilistic analysis, functional analysis, and assignment of annotation terms.

Data preparation

Data was initially organized into two datasets: one included all observations (gene array data) from samples at screening and the other used observations from LDRT samples. "Differenced" datasets were then created by comparing differential gene expression between screening and LDRT samples. In total there were six dataset partitions:

- All screening observations.
- All LDRT observations.
- Differenced observations for pooled placebo, responders and non-responders.
- Differenced non-responders and placebo.
- Differenced non-responders and responders.
- Differenced responders and placebo.

The 'All screening' subset was used as a baseline measure for the other five partitions in the predictive models. The lack of significant overlap of the 'All screening' dataset with the other groups indicated a generalized change in gene expression following therapy.

Gene selection

Hill climbing with a backtracking facility was used to search the data subsets and identify genes of predictive interest.[10] We compared possible subsets of genes and calculated the individual predictive ability of each gene to fit into a category (placebo, responder, non-responder) and corrected for correlations between genes. A ranking algorithm was then used to select gene subsets that had a high predictive ability and low intra-set correlation. Most genes were not predictive of categories (for example, housekeeping genes) and were discarded.

Bayesian network analysis

Bayesian network structure and parameter estimates[11] were learned directly from study generated data partitions. The program proceeded by assuming an equal likelihood for all models, sequentially searching the set of all models and then assigning a network score based on either Bayes Factor (BF) or Akaike Information Criterion (AIC) criteria. A heuristic search algorithm (K2 algorithm) was used to traverse the model space.[12]

Since both gene expression levels (absolute) and changes in gene expression (differenced) could be predictive of SCV-07 response, we assessed both. Using gene expression data from the following dataset partitions, we generated Bayesian networks based on four models:

1. *Screening data:* This model only considered absolute gene expression levels and was generated as a baseline test.
2. *LDRT data:* This model also used only absolute gene expression levels and was generated to determine if a single post-treatment set of gene expression data could accurately predict if a subject received SCV-07, received SCV-07 and responded clinically, or received SCV-07, but did not respond clinically.
3. Differenced screening data evaluating changes in gene expression. The network associated with the differenced screening data was generated to see if pre-treatment gene expression data, *combined* with information about the change in expression status between screening and LDRT, could accurately predict if a subject received SCV-07, received SCV-07 and responded well, or received SCV-07 and did not respond.
4. Differenced LDRT data, generated to see if post-treatment gene expression data, *combined* with information about the change in expression status between screening and LDRT, could accurately predict if a subject received no SCV-07, SCV-07 and responded well, or SCV-07 and did not respond.

Functional analysis

The gene list generated by the selection process was inspected for functional annotations using ontology guidelines, including 5213 molecular function, 1212 cellular component, and 7323 biological process categories.[9] Functional annotation data were queried and two types of analyses, a "baseline analysis" and a "functional annotation analysis," led to identification of candidate genes including search key identification names, gene names, and gene descriptions.

The *baseline analysis* recognized that all categorized genes were associated with functional annotations, assuming that genes sharing a similar set of annotations were more likely to be involved in related biological mechanisms. The purpose of this analysis was to identify clusters of similar genes that shared a set of terms, using the Kappa score as a measure of gene similarity.

The *functional annotation analysis* determined the functional relevance of gene clusters that were identified in the baseline analysis. The functional annotation analysis qualified, quantified and ranked the strength of the statistical relationship within the list of identified genes.

Combining Bayesian networks with functional analysis

To determine which genes were critical to the predictive value of the Bayesian network models, the clusters found in the baseline and functional annotation were mapped onto the LDRT Bayesian network using the BF score type and analyzed for functional categories. We examined the importance of these genes in defining responders and non-responders by re-evaluating network prediction accuracy with these genes removed.

Results

From the initial set of 26,100 genes, we first filtered out all genes that were identically expressed in all samples. We then designed an algorithm to identify differentially expressed genes in the categories of placebo, non-responder or responder. Screening samples were considered independently, and from this group 51 genes were identified by the algorithm and were used to generate a baseline. From the remaining five analytical groups 107 genes were identified and were used to build predictive models. Of the 107 genes, 68 have been characterized and named, based on prior experimental studies or homology to other genes.

Table 3
Network predictive accuracy and ROC results are shown broken out by the dataset used in the analysis and the Bayesian network score type used, with sample size in parenthesis. Bayes Factor is the odds of the marginal likelihood of one model compared to another model. AIC or Akaike Information Criterion is a measure of the goodness of fit of an estimated statistical model. The AIC score penalizes a model for each factor and accordingly is used to prevent model over-fitting.

Data (3-way)	Bayes Factor (BF)		Akaike Information Criterion	
Bayesian network score type				
All screening	Acc = 0.21 (24)	ROC = 0.35	Acc = 0.46 (24)	ROC = 0.60
All LDRT	Acc = 0.81 (27)	ROC = 0.89	Acc = 0.96 (27)	ROC = 0.99
Diff screening	Acc = 0.78 (23)	ROC = 0.84	Acc = 0.86 (23)	ROC = 0.95
Diff LDRT	Acc = 0.83 (23)	ROC = 0.99	Acc = 1.00 (23)	ROC = 1.00

Bayesian network analysis was performed using the four different models described above, each by two scoring methods, and generated a total of eight networks. The prediction accuracy (Acc) and sensitivity as described by area under the receiver operating characteristic (ROC) were determined for the three-way analyses (placebo vs. responders vs. non-responders) which allowed for a comparison between placebo and treatment, and also for the 'influence' of the placebo group to be 'subtracted out' from the responders and non-responders comparisons (Table 3).

A more specific analysis was performed to compare the predictive value of both the BF and AIC Bayesian network scores against a pure guess (Table 4). Both analyses demonstrated the 'random' state of the placebo group vs. the treatment group and the responders group vs. non-responders group based on screening data. With screening data only, even the best-fitting Bayesian network could not predict if a subject would be later assigned to the placebo group or the treatment group. With the exception of the screening data networks, the results show the superiority of the AIC score type over the BF score type in more accurately differentiating between placebo and treatment groups. Both score types did well in discriminating between responders and non-responders.

The baseline analysis resulted in two clusters of genes. These were functionally categorized and found to be associated with responder/non-responder status with $p < 0.10$ (Supplemental Tables 1s and 2s), via goodness of fit of logistic regression based on gene expression.[10]

Table 4
Predictive accuracy results are shown broken out by the dataset used in the analysis, the Bayesian network score type used, and the objective that the two way network accomplishes. Counts show the number of subjects that were used to create each network, accuracy refers to the predictive accuracy of the Bayesian network, pure guess refers to the predictive power of a pure guess, and delta is calculated as the difference between the accuracy and the pure guess values, rounded.

Dataset	Count	Placebo vs. treatment			Responders vs. non-responders		
		Accuracy	pure guess	delta	Accuracy	pure guess	delta
Bayes Factor Bayesian network score type							
All screening	24	58%	67%	−8%	25%	56%	−31%
All LDRT	27	85%	70%	15%	94%	53%	41%
Diff screening	23	78%	70%	9%	100%	56%	44%
Diff LDRT	23	91%	70%	22%	93%	56%	37%
AIC Bayesian network score type							
All screening	24	67%	67%	0%	50%	56%	−6%
All LDRT	27	96%	70%	26%	100%	53%	47%
Diff screening	23	91%	70%	22%	93%	56%	37%
Diff LDRT	23	100%	70%	30%	100%	56%	44%

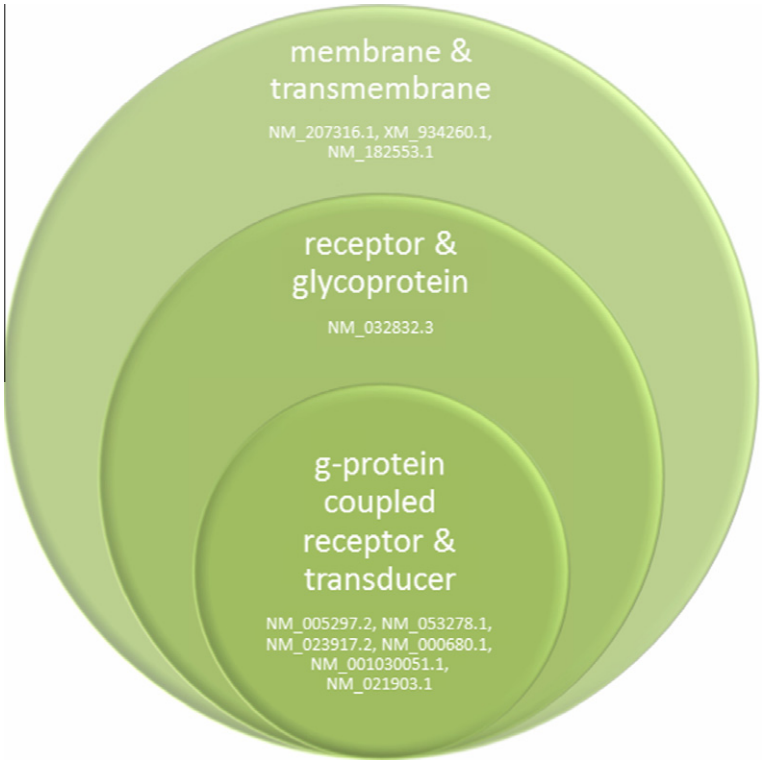

Figure 1 Stacked Venn diagram showing genes grouped by functional annotations. Genes are represented in a functional annotation term group if they are contained within the circle associated with the name or in any smaller (stacked).

A cluster of 10 genes with statistically significant functional annotation terms were defined by the functional annotation analysis (see Supplemental Tables 1s and 2s). The predictive importance of the 10 genes in the functional annotation cluster was tested by seeing how removal of these genes affected the model. The original model featured a predictive accuracy of 81% (27 observations), while the modified network model (removing the cluster of 10 genes) had a predictive accuracy of only 70%. The area under the ROC of the original network was 89%. By convention, area under the ROC prediction accuracies $\geqslant$80% are defined as 'Good' and $\geqslant$90% is 'Excellent'.[13]

We inspected the 10 genes for the strength of association with the functional annotations. The significance of the relationship of these genes to the functional annotations was determined, after multiple hypothesis correction (Benjamini–Hochberg), and was, based on the strength of the association and their functional annotations, highly significant. The relationship between the 10 genes and each of their functional annotations is shown in a stacked Venn diagram (Fig. 1).

Discussion

We determined if differences in gene expression could differentiate individuals' SCV-07 responsiveness leading to a delay in OM. Our approach was based on the hypothesis that differences in SCV-07 responsiveness were more likely to be discriminated by the simultaneous expression of multiple genes (a cluster), then by one or two individual genes. The analysis was undirected and did not assume that a differentiation between responders and non-responders was necessarily associated with SCV-07's biological activity. Models were 'learned' from the expressed genes and resulted from the accumulated knowledge generated by sequentially comparing multiple gene combinations unique to the categories defined by the study.

We utilized RNA obtained at screening and LDRT from study subjects who received placebo or SCV-07 and who were classified clinically as responders or non-responders. As expected, the vast majority of expressed genes were common to all study cohorts, largely comprised of housekeeping genes, and were not included. The randomness of cohort assignment was confirmed, as screening data were unable to predict study group assignment.

We were able to discriminate between subjects assigned to placebo or SCV-07 treatment based on the expression of 107 genes, 68 of which had functional characterization previously described in the literature. Additional analysis is necessary to determine the potential mechanistic import of each gene, and presents opportunities to characterize some or all of the genes with unknown functional definitions.

We evaluated these 107 genes associated with SCV-07 administration, to discriminate SCV-07 responders from non-responders using four models. The LDRT and two differenced datasets predicted treatment from placebo and responder vs. non-responder status with good accuracy. In the two Bayesian network tests of predictive accuracy, AIC were superior to BF: the accuracy of predicting placebo vs. treatment for the differenced screening dataset was 78% using the BF, but 91% using the AIC score. Our ability to differentiate responders from non-responders using the AIC score was dramatic: 100% vs. a guess of 53% for the LDRT dataset, 93% vs. 56% for the differenced screening dataset, and 100% vs. 56% for the differenced LDRT dataset.

The analysis produced eight predictive Bayesian networks based on different input datasets. Functional cluster analysis resulted in finding six associations that reached statistical significance. A 10 gene cluster was identified that would have been unlikely to occur at random from the human genome. This set of genes was associated with the terms g-protein coupled receptor, transducer, receptor, transmembrane, glycoprotein, and membrane, indicating that the Bayesian networks learned in the

analysis may have isolated not only a set of statistical dependencies, but also biological dependencies. Elimination of this cluster abrogated the predictability of the dataset.

Our ability to associate gene expression changes with SCV-07's mechanism of action was limited by sample collection timing. SCV-07's maximum clinical benefit was seen from the initiation of radiotherapy until a cumulative radiation dose of 45 Gy, so it is likely that analysis of specimens obtained at time points earlier in radiation treatment would be more useful to establish the mechanism(s) by which SCV-07 attenuates OM.

In the context of a clinical trial, the value of our results falls into two major categories. First, the finding that SCV-07-treated subjects expressed a gene set that were not expressed by placebo patients indicates a biological response to the drug. This conclusion is supported by the observation of differences in blood cytokine levels noted between placebo- and SCV-07-treated patients.[4] Second, the observation that a specific gene cluster was seen in SCV-07 responders but not in non-responders suggests that identifiable genetic markers could be used prospectively to determine the best candidates for SCV-07 treatment. Since our predictive models required post-treatment data, the translation of our results to a predictive test for SCV-07 responsiveness requires additional studies.

Personalized medicine provides advantages to patients, clinicians, and the healthcare system and has the potential to direct medications to patients who are most likely to benefit. We report here, for the first time, the application of a genomics-based analytical paradigm that demonstrates proof of concept for discriminating responders and non-responders for an OM intervention.

Conflict of interest statement

Dr. Alterovitz serves as a consultant to Biomodels, LLC. Drs. Tuthill, Rios and Modelska are pain employees of SciClone Pharmaceuticals. Dr. Sonis is a partner in Biomodels, LLC.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.oraloncology.2011.07.006.

References

1. Sonis S. Mucositis: the impact, biology and therapeutic opportunities of oral mucositis. *Oral Oncol* 2010;**45**:1015–20.
2. Watkins B, Watkins B, Pouliot K, et al. Attenuation of radiation- and chemothoradiation-induced mucositis using gamma-D-glutamyl-L-tryptophan (SCV-07). *Oral Dis* 2010;**16**:655–60.
3. Adkins D et al. Safety and efficacy of gamma-D-glutamyl-L-tryptophan as an intervention for oral mucositis in patients receiving chemoradiation for the treatment of cancers of the head and neck. *J Clin Oncol* 2010;**28**(suppl abstr):e19693.
4. Adkins D, et al. SCV-07 as an intervention for oral mucositis in head and neck cancer patients receiving chemo-radiation therapy – Phase 2a trial results. International Society for Biological Therapy Cancer 25th Annual Meeting; 2010.
5. Bansal M, Belcastro V, Ambesi-Impiombata A, et al. How to infer gene networks from expression profiles. *Mol Syst Biol* 2007;**3**:78.
6. Needham CJ, Manfield IW, Bulpitt AJ, et al. From gene expression to regulatory networks in *Arabidopsis thaliana*. *BMC Syst Biol* 2009;**3**:85–102.
7. Salonga D, Danenberg KD, Johnson M, et al. Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clin Cancer Res* 2000;**6**:1322–7.
8. Padyukov L, Lampa J, Heimburger M, et al. Genetic markers for efficacy of tumor necrosis factor blocking therapy in rheumatoid arthritis. *Ann Rheum Dis* 2003;**62**:526–9.
9. Sonis ST, Elting LS, Keefe D, et al. Perspectives on cancer therapy-induced mucosal injury. *Cancer* 2004;**100**:1995–2025.
10. Lemeshow Stanley, Hosmer Jr David W. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;**115**(1):92–106.
11. Wilkinson D. Bayesian methods in bioinformatics and computational systems biology. *Brief Bioinform* 2007;**8**:109–16.
12. Markowetz F, Sprang R. Inferring cellular networks – A review. *BMC Bioinform* 2007;**8**(suppl 6):S5.
13. Jesse M, Pines J, Everett W. *Evidence-based emergency care: diagnostic testing and clinical decision rules*. Blackwell Publishing; 2008.