# Statistical Analysis Plan

## An 8-Week Prospective Randomized, Controlled, Double-Blind Trial of the Genecept Assay™ vs. Treatment-as-Usual to Evaluate Efficacy of Assay-Guided Treatment in Adults with Major Depressive Disorder

| | |
|---|---|
| Protocol ID: | GNM-PROT MDD-01 |
| Product: | Genecept Assay™ |
| Version: | **Version 4.0** |
| Date: | 30-Jun-2017 |
| Previous Version | Version 3.0, dated 28-Nov-2016 |
| Study Sponsor: | Genomind, Inc.<br>2200 Renaissance Blvd., Ste. 100<br>King of Prussia, PA 19406<br>1-877-895-8658 |

**Confidential:** This document contains confidential information that is the property of Genomind, Inc. and must not be disclosed to anyone other than the recipient clinical Investigator(s), their designees, and members of the Institutional Review Board. This information must not be used for any purpose other than the evaluation or conduct of the clinical investigation without the prior written consent of Genomind,

# Statistical Analysis Plan Approval Page

## Version 3.0

An 8-*Week* Prospective Randomized, Controlled, Double-Blind Trial of the Genecept Assay™ vs. Treatment-as-Usual to Evaluate Efficacy of Assay-Guided Treatment in Adults with Major Depressive Disorder

Protocol ID: GNM-PROT MDD-01

Product: Genecept Assay™

Sponsor: Genomind, Inc.

| | |
|---|---|
| Signature: | Signature: |
| Date: | Date: 5 July 2017 |
| Larry Roi, PhD<br>Study Statistician<br>BDM Consulting, Inc.<br>100 Franklin Square Drive, Suite 102<br>Somerset, NJ 08873 | Kirsti Cook, MA<br>Director, Clinical Operations<br>Genomind, Inc<br>2200 Renaissance Blvd., Suite 100<br>King of Prussia, PA 19406 |


Confidential Property of Genomind, Inc.

# 0 ABBREVIATIONS

| | |
|---|---|
| AGT | Assay Guided Treatment |
| ANCOVA | Analysis of Covariance |
| ATRQ | Antidepressant Treatment Response Questionnaire |
| CBT | Cognitive Behavioral Therapy |
| CGI-I | Clinical Global Impression - Improvement |
| CGI-S | Clinical Global Impression - Severity |
| CSR | Clinical Study Report |
| C-SSRS | Columbia Suicide Severity Rating Scale |
| DBT | Dialectical Behavioral Therapy |
| ECT | Electroconvulsive Therapy |
| eCRF | Electronic Case Report Form |
| FAS | Full Analysis Set |
| FIBSER | Frequency, Intensity, and Burden of Side Effects Ratings |
| SIGH-D-17 | Structured Interview Guide for the Hamilton Depression Rating Scale – 17 Item Version |
| ICH | International Conference on Harmonization |
| IWRS | Interactive Web Response System |
| LOCF | Last Observation Carried Forward |
| LS | Least Squares |
| MAR | Missing at Random |
| MDD | Major Depressive Disorder |
| MINI | M.I.N.I. International Neuropsychiatric Interview |
| MMRM | Mixed Model Repeated Measures |
| NMAR | Not Missing at Random |
| OC | Observed Cases Analysis Set |
| PRISE | Patient Rated Inventory of Side Effects |
| QIDS-SR16 | Quick Inventory of Depressive Symptomatology-Self Report (16 item) |
| SD | Standard Deviation |
| SAF | Safety Analysis Set |
| SAP | Statistical Analysis Plan |
| TAU | Treatment as Usual |
| TEAE | Treatment Emergent Adverse Event |
| TMS | Transcranial Magnetic Stimulation |
| TRD | Treatment Resistant Depression |


Confidential Property of Genomind, Inc.

# 1 OVERVIEW AND STUDY PLAN

This Statistical Analysis Plan (SAP) provides a prospective plan for data handling and statistical analyses for the Genomind MDD-001 study "An 8-Week Prospective Randomized, Controlled, Double-Blind Trial of the Genecept Assay ™ vs. Treatment-as-Usual to Evaluate Efficacy of Assay-Guided Treatment in Adults with Major Depressive Disorder." The SAP adds detail to the statistical methods given in the study protocol, and ensures credibility of study findings by pre-specifying the key statistical approaches prior to study start. As, necessary, SAP amendments clarify and document changes to statistical procedures, or align the SAP with protocol amendments.

## 1.1 SAP Versions

**SAP Version 2.0** detailed the analyses for the main study (n ≅ 300, adult MDD). It was finalized, signed, and approved by the Sponsor before the first main study subject was randomized**.**

**SAP Version 3.0** incorporated Amendment 2 changes to the protocol and made minor technical corrections to the main study analyses. Protocol Amendment 2 added a sub-study, the Exploratory Elderly MDD Study, a follow-on to main study of ≅ 70 subjects, age ≥ 65, utilizing a subset of sites in the main study. SAP Version 3.0 was signed and approved by the Sponsor before the main study was unblinded.

**SAP Version 4.0**, this version, aligns the SAP with Protocol Amendment 3, which added details regarding the Exploratory Elderly MDD Study, and clarified that this study would be analyzed separately from the main study. SAP Version 4.0, makes no changes to the main study analyses, but rather makes changes to the Exploratory Elderly MDD Study analyses consistent with its exploratory nature

## 1.2 Study Description

This study compares efficacy and safety outcomes in Major Depressive Disorder (MDD) adult patients randomized to assay-guided treatment (AGT) or treatment-as-usual (TAU). The treatment duration will be 8-weeks. Subjects will be assessed at visits at Week 2, 4, 6 and 8. Approximately 300 subjects will be randomized 1:1 to the two treatment group (AGT and TAU). This is a multi-center trial, with approximately 25 sites in the US.

Randomization will be by IWRS. The treating investigator will be unblinded to treatment assignment (necessarily). Other site staff, sponsor staff (including site monitors) and all others will be blinded to treatment assignment for the duration of the subject's participation in the study. The (blinded) rater for the primary endpoint, the SIGH-D-17 Hamilton Depression Scale, will have no other contact with the subject such as collection of screening data, follow-up assessments, documentation of adverse events, etc. Blinded raters will not discuss subjects with other study staff.

After recruitment for main study is completed, an additional 70 subjects , age 65 years and older will be randomized to the Exploratory Elderly MDD Study. This follow-on sub-study will apply all procedures of the main study to this elderly population subset.


Confidential Property of Genomind, Inc.

## 1.3 Study Objectives

### Primary Objective

- **To assess efficacy of assay-guided treatment (AGT) versus treatment-as-usual (TAU) in Major Depressive Disorder, as measured by change in Hamilton Depression Rating Scale (SIGH-D-17) at 8 weeks.**

### Secondary Objectives

- **To assess efficacy of assay-guided treatment (AGT) versus treatment-as-usual (TAU) in Major Depressive Disorder, as measured by change in Quick Inventory of Depressive Symptoms (QIDS-SR16) at 8 weeks.**
- **To assess the percentage of responders at Week 8, based on SIGH-D-17, QIDS-SR16, and Clinical Global Impression–Improvement (CGI-I), respectively. Treatment response will be defined in 3 different ways:**
  - Blinded Rater Assessment: ≥ 50% reduction from baseline of SIGH-D-17 score ;
  - Subject-Rated Assessment: ≥ 50% reduction from baseline of QIDS-SR16 ;
  - Unblinded Clinician Assessment: ≤ 3 score on the CGI-I;
- **To assess the percentage of remitters at Week 8, based on the SIGH-D-17 and QIDS-SR16, respectively. Remission is defined for the SIGH-D-17 and the QIDS-SR16:**
  - Blinded Rater Assessment: ≤ 7 score on the SIGH-D-17;
  - Subject-Rated Assessment: ≤ 5 score on the QIDS-SR16
- **To assess the impact of the Genecept Assay ™ on adverse events, based on Frequency, Intensity and Burden of Side Effects Rating (FIBSER) and Patient-Related Inventory of Side Effects (PRISE), frequency and severity of adverse events and Columbia-Suicide Severity Rating Scale (C-SSRS) outcomes.**
- After completion of the main MDD study, to conduct the Exploratory Elderly MDD Study, to preliminarily assess the efficacy and safety of AGT vs TAU in the elderly, age ≥ 65. This study will provide additional data on this important, Medicare-eligible population, which is sparsely represented in the Main Study (< 8% of subjects). As specified in Protocol Amendment 2, after the Main Study enrollment finishes, approximately 70 subjects, age ≥ 65, will be randomized to the this add-on study. Except for change in age of inclusion, all study procedures will be identical to the Main Study. Primary efficacy will be measured by change in Hamilton Depression Rating Scale (SIGH-D-17) at 8 weeks.


Confidential Property of Genomind, Inc.

## 1.4 Randomization

Approximately 300 for the main study, plus an additional 70 subjects age ≥ 65 for the Exploratory Elderly MDD Study, will be randomized in a 1:1 ratio to receive Assay-Guided Treatment (AGT) or Treatment as Usual (TAU). Randomization will be stratified by site, using a variable block size randomization to minimize bias. Randomization will be completed via IWRS during the Baseline visit, after Screening and Baseline inclusion/exclusion criteria have been satisfied.

# 2 STATISTICAL AND ANALYTICAL PROCEDURES

**Note:** *Unless stated otherwise, this section applies to both the Main Study and the Exploratory Elderly MDD Study. Procedures, in general, may be adjusted in the Exploratory Elderly MDD Study to reflect sparse sample sizes, eg, substitution of a short listing in place of a full table.*

Summarizations will be provided in tables and figures. Safety and efficacy variables will be summarized by visit using descriptive statistics. For continuous variables, this will include the number of non-missing values, mean, standard deviation (SD) or standard error of the mean, median, minimum, and maximum. Categorical variables will be summarized using counts and percentages.

Unless otherwise stated, the significance level for statistical tests will be two-sided, alpha=0.05, with no adjustment for multiple comparisons

Pooled Investigative Sites: For all statistical analyses, sites with few subjects (< 7 subjects) will be pooled into a single larger site. "Investigative site" in all statistical analyses will refer to "sites" as pooled.

SIGH-D-17 interviews will be reviewed by central expert monitors, and site SIGH-D-17 scores may be revised based on the central monitor evaluation. All SIGH-D-17 changes due to central review will be described by listings. SIGH-D-17 statistical analyses, summaries, and listings will otherwise rely solely on the revised SIGH-D-17 scores. Note, SIGH-D-17 inclusion criteria are based on the original site rater screening and baseline SIGH-D-17, irrespective of any later revision after central review.

**Note:** *No central rater evaluation will be done for the Exploratory Elderly MDD Study, as the sites enrolling subjects will continue from the Main Study and already have trained, experienced site raters.*

Data will be reported and analyzed using SAS software, version 9.3, except as noted otherwise.

## 2.1 Efficacy Analysis Estimands

This trial is a comparison of outcomes between two treatment strategies, evaluated at 8 weeks. For practical interpretation, this implies the statistical analyses should estimate and test the difference in Week 8 outcomes for all subjects, irrespective of attrition or compliance. [This may also be described as following the Intent to Treat (ITT) principle, but there is confusion about the use of this term when there is missing data, and ITT terminology will not be used further here.]. Specifying the "estimands") (what is being estimated), aligned with the study purpose, is the basis for the formal statistical methods which follow.

### Primary Efficacy: AGT vs TAU

- -Difference at Week 8 in SIGH-D-17 mean change from baseline, for all randomized subjects.

### Secondary Efficacy: AGT vs TAU

- Difference at Week 8 in QIDS-SR16 mean change from baseline, for all randomized subjects.
- Difference in % treatment responders at Week 8, for all randomized subjects. Treatment response will be defined in 3 different ways:

  ≥ 50% reduction of SIGH-D-17 from baseline for blinded rater assessment

  ≥ 50% reduction of QIDS-SR16 from baseline for subject-rated assessment

  Score of ≤ 3 on the CGI-I from baseline for the unblinded clinician assessment
- Difference in % treatment remitters at Week 8, for all randomized subjects. Remission is defined for the SIGH-D-17 and the QIDS-SR16:

  Score of ≤ 7 on the SIGH-D-17 from baseline for blinded rater assessment

  Score of ≤ 5 on the QIDS-SR16 from baseline for subject-rated assessment.

### Other Efficacy

- Difference at Week 8 in mean % improvement from baseline in SIGH-D-17 scores, for all randomized subjects.
- Difference at Week 8 in mean % improvement from baseline in QIDS-SR16 scores, for all randomized subjects.
- Difference at Week 8 in CGI-I % scores, for all randomized subjects.
- Difference at Week 8 in CGI-S Change from baseline scores, for all randomized subjects.

## 2.2 Safety Endpoints

- Treatment Emergent Adverse Events (TEAEs)
- Frequency and severity of medication-induced side effects based on collection of FIBSER, PRISE and Adverse Events
- Columbia-Suicide Severity Rating Scale (C-SSRS) Change from baseline.


Confidential Property of Genomind, Inc.

## 2.3 Other Endpoints

- Clinical Decision Survey Findings
- Pre-Result: Clinician intended treatment prior to receiving assay results
- Post-Result (AGT only): Clinician rating of confidence in making treatment decisions, treatment plan and gene results which influenced treatment plan after results were received.

## 2.4 Analysis Datasets

**All Randomized Set**: All randomized subjects. Special handling of patients entering the trial more than once: Only the data for the first study participation will be analyzed. Data from the second study participation will be reviewed for safety, but not included in any analysis set or analysis.

**Full Analysis Set (FAS)**: All randomized subjects with baseline and post-baseline SIGH-D-17 assessment. The subject who was treated before randomization (116-001) will not be included in the FAS.

**Safety Analysis Set (SAF)**: All randomized subjects who complete the (post-randomization) baseline appointment with the treating investigator, i.e., start AGT or TAU treatment.

The **FAS** will be the primary dataset for all efficacy analyses. Note that, as described above, the desired estimands are based on all randomized patients. It is anticipated the percent of patients missing post-baseline SIGH-D-17 will be small, and thus the FAS analyses are valid, and may be done with fewer statistical assumptions.

The **All Randomized Set** data will be listed, and the reason for dropout of patients without post-baseline, will particularly be reviewed. If the number of FAS subjects is less than 95% of the All Randomized Set, the primary and secondary analyses will be conducted on this dataset as sensitivity analyses.

# 3 STATISTICAL METHODS

**Note:** *Unless stated otherwise, this section applies to both the Main Study and the Exploratory Elderly MDD Study. In general, in the Exploratory Elderly MDD Study analyses may be adjusted to reflect sparse sample sizes, e.g., substitution of a short listing in place of a full table.*

*The efficacy analyses for the Exploratory Elderly MDD Study are specified at the end of the Efficacy* ***Analyses, Section 3.4****.*

## 3.1 Subject Disposition

The number of subjects will be summarized by treatment group and study center for the three analysis datasets. Randomized subjects excluded from the SAF and FAS populations will be listed. The number and percentage of subjects per treatment group who prematurely discontinued treatment will be presented by visit and overall, by treatment group, for SAF population. The number and percentage of subjects per treatment group who prematurely discontinued from the study will be presented by visit and overall, for the SAF and FAS populations. A CONSORT style (Ref) flow chart will describe premature study discontinuation for the FAS population. The coded Reason for Premature Discontinuation will be summarized (number and percentage) by treatment group for the FAS Population. The verbatim Reason for Discontinuation and Additional Information text entries


Confidential Property of Genomind, Inc.

will be listed. The coded Reason for Discontinuation and verbatim Reason for Discontinuation and Additional Information may be used in sensitivity analyses for the primary efficacy analysis.

## 3.2 Demographic and Pretreatment Characteristics

Demographics (age, race, sex, weight, height, and BMI), disease severity [including SIGH-D-17 at screening and baseline; QIDS-SR16 and CGI-S at baseline], number of failed, adequate treatment regimens in the current MDD episode (1 vs > 1), previous depression treatment/history, and other screening/baseline visit assessments will be summarized by visit (Screening, Baseline) and treatment group using descriptive statistics. The number of failed, adequate treatments will be considered to be > 1 if two or more standard MDD medications (FDA approved for MDD or approved by the sponsor as commonly used for MDD) are entered in the eCRF as "ATRQ Failed Medications," with daily dose greater than or equal to the minimum MDD dose, and given for at least 6 weeks.

## 3.3 Depression Treatment and Concomitant Medications

### Depression Drug Treatment

On-study depression drugs (taken between Day 0 and Week 8 visit) will be summarized by the number and proportion of subjects in each treatment group receiving each medication within each drug class, for the Safety Population. Treatment Exposure (days) to on-study depression drugs (days) from Day 0 to Week 8 will be summarized for the Safety Population by treatment group and drug class. Descriptive statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented.

### Concomitant Medications

CNS drugs (taken between Day 0 and Week 8 visit) will be summarized by the number and proportion of subjects in each treatment group receiving each medication within each drug class, for the Safety Population. Multiple medication use by a subject will be counted only once.

All other concomitant medications use (medication taken between Day 0 and Week 8 visit) will be listed. Additional summarizations may be done to help interpret study findings.

## 3.4 Efficacy Analyses

**Note**: *The Efficacy Analyses for the Exploratory Elderly MDD Study will be given at the end of this section. Generally, the statistical methods for the Exploratory Elderly MDD Study will follow the outline of the Main Study, but will be primarily descriptive, and statistical analyses will be simpler, as necessitated by the smaller sample size.*

### Primary Endpoint Analyses

A Mixed Model Repeated Measures (MMRM) analysis will test whether AGT shows greater improvement in SIGH-D-17 at Week 8 than TAU. Formally, the analysis tests for a difference in mean change from baseline to Week 8 between the treatment groups, in the FAS population. Compared to analyzing All Randomized subjects, this excludes subjects who don't return after the Baseline Visit, which is generally acceptable, if this excluded group is small.

In the MMRM model, the dependent variable is the change in SIGH-D-17 from baseline at each of the scheduled visits Week 2, Week 4, Week 6 and Week 8. The model will include the fixed effect continuous factor baseline SIGH-D-17, and fixed effect categorical factors investigative site, treatment group (AGT and TAU; 2 levels), visit (Weeks 2, 4, 6 and 8; 4 levels), and treatment x visit interaction. The interaction term will remain in the model regardless of statistical significance. An unstructured covariance pattern will be used to estimate the variance-covariance matrix of the within-subject repeated measures (i.e., model within-subject errors). The variance-covariance matrix will be estimated across both treatment arms. The fixed effects will be estimated by the method of restricted maximum likelihood estimation using the Newton-Raphson algorithm, based on all available data. The denominator degrees of freedom and adjustment of standard errors will use the Kenward-Roger method.

Assignment of SIGH-D-17 assessments to Visit for primary MMRM analysis:

Day 7-20 SIGH-D-17 => Week 2 Visit
Day 21-34 SIGH-D-17 => Week 4 Visit
Day 35-48 SIGH-D-17 => Week 6 Visit
Day 49-70 SIGH-D-17 => Week 8 Visit

If the final SIGH-D-17 assessment is on Day 1-6, this SIGH-D-17 value will be assigned to the Week 2 Visit. If multiple SIGH-D-17 assessments fall within a Visit assignment range, the assessment closest in time to the nominal visit day will be retained. For example, the in range assessment taken closest to Day 14 for Week 2 Visit, will be retained.

AGT and TAU mean change in SIGH-D-17 at Week 8 will be estimated and tested utilizing the least squares (LS) means from the treatment x visit interaction in the MMRM model. The primary analysis will test the difference (contrast) between the Week 8 LS means, at two-sided significance 0.05. The test p-value and the 95% CI for the difference in Week 8 means will be presented. Comparisons between the AGT and TAU means at Week 2, Week 4 and Week 6 will also be generated. Diagnostics for outliers, non-normality and other model assumptions will be run. Findings will be followed up with sensitivity analyses, as appropriate.

To allow comparison with previous SIGH-D-17 studies, two supportive analyses of the primary endpoint will be produced. Each will be an Analysis of Covariance (ANCOVA):

1) Last Observation Carried Forward (LOCF) analysis, and

2) Observed Cases (OC) analysis.

Each ANCOVA will include baseline SIGH-D-17, Investigative Site and Treatment Group as the independent variables in the model. The dependent variable for the LOCF analysis will be last available SIGH-D-17 change from baseline. The dependent variable for the OC analysis will be the Week 8 SIGH-D-17 score, with no imputation of missing Week 8 scores.

## Missing Data

The MMRM analysis may be biased unless the missing SIGH-D-17 visit data are Missing at Random (MAR). This is because the MMRM, as well as LOCF and OC ANCOVA analyses assume the unobserved missing data has the same behavior as the observed data for the same treatment group and covariates. [For example, if a computer failure causes loss of some Week 8 SIGH-D-17 scores, this missing data should be MAR: the missing data is not informative about the actual Week 8 SIGH-D-17


Confidential Property of Genomind, Inc.

scores, and the missing scores may be predicted from available SIGH-D-17 scores. Contrarily, if scores are missing due to subject dropout resulting from , (un-measured) worsening depression, then these missing scores would likely be Not Missing at Random (NMAR)].

To address this issue, if there is more than 10% missing Week 8 SIGH-D-17 scores, sensitivity analyses using alternate approaches to handling missing data will be explored. The sensitivity analyses will incorporate a range of plausible NMAR assumptions. As methods for NMAR sensitivity analyses are currently evolving, further technical detail will be given in a later amendment to the SAP.

## Subgroup Analyses

Consistency of AGT vs TAU treatment effect on SIGH-D-17 will be assessed across various subgroup factors described below, by evaluating the treatment by subgroup interaction in the MMRM model. If there is evidence of inconsistency of treatment effect (i.e., p-value <0.10 for interaction test), further analyses may be performed. LS Mean Differences for each subgroup factor will be graphically displayed.

| **Subgroup factor** | **Subgroup categorization** |
|---|---|
| Age group | < median age, ≥ median age |
| Gender | Male, Female |
| Race | White, Non-white |
| Baseline SIGH-D-17* | < 24, ≥ 24 (Moderate, Severe MDD) |
| Previous failed drug treatments | 1, >1 Previous failed, adequate trials of antidepressants in the current depressive episode |
| Investigative Site | Pooled sites. |

*The continuous baseline SIGH-D-17 term will not be in the MMRM model.

Graphical presentations of the primary efficacy subgroup results will also be provided

## Analyses of Key Secondary Efficacy Variables

The test for the following secondary estimates based on continuous variables will use the MMRM model in the FAS population, as described for the primary analysis:

- Difference at Week 8 in QIDS-SR16 mean change from baseline.
- Difference at Week 8 in mean % improvement from baseline in SIGH-D-17 scores.
- Difference at Week 8 in mean % improvement from baseline in QIDS-SR16 scores.


Confidential Property of Genomind, Inc.

- Difference at Week 8 in mean change from baseline of CGI-S.

The Mantel-Haenszel method, stratified by investigative site, in the FAS population, will test AGT vs TAU estimates based on categorical variables:

- Difference in % treatment responders at Week 8, done separately for SIGH-D-17, QIDS-SR16 and CGI-I. Treatment response is defined as ≥ 50% reduction of SIGH-D-17 from baseline to Week 8 for blinded rater assessment, ≥ 50% reduction of QIDS-SR16 from baseline to Week 8 for subject-rated assessment, score of ≤ 3 on the CGI-I from baseline to Week 8 for the unblinded clinician assessment. If Treatment Response is missing at Week 8, the subject will be considered a Week 8 treatment non-responder. The two-sided p-value, estimated odds ratio and 95% CI for the odds ratio will be computed.

- Difference in % treatment remitters, done separately for SIGH-D-17 and QIDS-SR16, at Week 8. Remission is defined as score of ≤ 7 on the SIGH-D-17 from baseline to Week 8 for blinded rater assessment, score of ≤ 5 on the QIDS-SR16 from baseline to Week 8 for subject-rated assessment. If Treatment Remission is missing at Week 8, the subject will considered a Week 8 treatment non-responder. The two-sided p-value, estimated odds ratio and 95% CI for the odds ratio will be computed

- Difference at Week 8 in CGI-I profile.

***==========***

## *Efficacy Analyses for Exploratory Elderly MDD Study*

### *Primary Analysis*

The Exploratory Elderly MDD Study primary analysis will test the Difference at Week 8 in mean SIGH-D-17 change from baseline in a Last Observation Carried Forward ANCOVA, adjusting for baseline SIGH-D-17 and pooled site, in the FAS set. Descriptive statistics and confidence intervals will also be calculated. The Treatment x Pooled Site and Treatment x Baseline SIGH-D-17 interaction tests, as well as diagnostics for outliers, non-normality and other model assumptions will be run, and followed up, as appropriate.

### *Subgroup Analyses*

Consistency of AGT vs TAU treatment effect on SIGH-D-17 will be assessed across subgroup factors described below, by evaluating the treatment by subgroup interaction in the LOCF ANCOVA model. If there is evidence of inconsistency of treatment effect (i.e., p-value <0.10 for interaction test), further analyses may be performed. Subgroups will include Gender, Baseline SIGH-D-17 (categorized as in the Main Study, continuous baseline SIGH-D-17 not in the model) and Pooled Site. Confidence intervals for the TAU minus AGT LS Mean Differences for each subgroup factor will be graphically displayed.


Confidential Property of Genomind, Inc.

***Analyses of Key Secondary Efficacy Variables***

The major emphasis will be on descriptive statistics and confidence intervals. Nominal p-values will also be computed.

The following secondary continuous variables will use the same LOCF ANCOVA model described for the primary analysis. Two-sided p-values, LS means and CIs for TAU - AGT difference at for Week 8 will be computed.

- Difference at Week 8 in QIDS-SR16 mean change from baseline.
- Difference at Week 8 in mean % improvement from baseline in SIGH-D-17 scores.
- Difference at Week 8 in mean % improvement from baseline in QIDS-SR16 scores.
- Difference at Week 8 in mean change from baseline of CGI-S.

Similarly, the Mantel-Haenszel method, stratified by investigative site, in the FAS population, will test AGT vs TAU based on categorical variables. The two-sided p-value, estimated odds ratio and 95% CI for the odds ratio will be computed:

- Difference in % treatment responders at Week 8, done separately for SIGH-D-17, QIDS-SR16 and CGI-I. Treatment response is defined as ≥ 50% reduction of SIGH-D-17 from baseline to Week 8 for blinded rater assessment, ≥ 50% reduction of QIDS-SR16 from baseline to Week 8 for subject-rated assessment, score of ≤ 3 on the CGI-I from baseline to Week 8 for the unblinded clinician assessment. If Treatment Response is missing at Week 8, the subject will be considered a Week 8 treatment non-responder.
- Difference in % treatment remitters, done separately for SIGH-D-17 and QIDS-SR16, at Week 8. Remission is defined as score of ≤ 7 on the SIGH-D-17 from baseline to Week 8 for blinded rater assessment, score of ≤ 5 on the QIDS-SR16 from baseline to Week 8 for subject-rated assessment. If Treatment Remission is missing at Week 8, the subject will considered a Week 8 treatment non-responder.
- Difference at Week 8 in CGI-I profile.

**End of Efficacy Analyses for the Exploratory Elderly MDD Study**

==========

## 3.5 Safety Analyses

**Note:** *This section applies to both the Main Study and the Exploratory Elderly MDD Study, although safety analyses may be adjusted in the Exploratory Elderly MDD Study to reflect sparse sample sizes, eg, substitution of a short listing in place of a full table.*

All safety analyses will be performed on the SAF dataset.

- Adverse events

The analysis will focus on the TEAEs. TEAEs will consist of all AEs with start date after Baseline. Continuing pre-Baseline AEs which worsen in severity or seriousness will be reported in the CRF as new AEs, and thus will be TEAEs.


Confidential Property of Genomind, Inc.

TEAEs will be summarized and tabulated according to body system and frequency for each treatment group. Separate tabulations of AEs depending on seriousness, severity, relationship to MDD treatment, action taken and outcome will be given. SAEs will be correspondingly summarized.

- FIBSER and PRISE

Overall and by visit results from the FIBSER and PRISE scales will be summarized descriptively by treatment.

### Other Analyses

All other endpoints will be descriptively summarized, including the Pre-Result and Post-Result Clinical Decision Survey Findings.

## 3.6 Interim Analysis

No interim analysis is planned for this study.

# 4 SAMPLE SIZE JUSTIFICATION

Sample size is calculated for the primary endpoint, improvement from baseline to eight weeks in the SIGH-D-17 of subjects with Major Depressive Disease.

As there are few studies comparing MDD AGT vs TAU, information to guide sample size is limited.

Nevertheless, three controlled MDD trials of AGT vs TAU[1-3] provide some quantitative basis for sample size. All are single center, 8-10 weeks in duration, measured SIGH-D-17, and utilized the AssureRx assay battery. For each study an approximate SIGH-D-17 Delta and SD were approximated from tables and figures in the report publication. Delta is the between group difference in mean SIGH-D-17 change and SD is the assumed constant standard deviation of the individual subject's change in HAMD). Cohen's d = Delta/SD was further calculated. Brief summaries of the three studies follow:

1) Hall-Flavin DK et al 2012, *Using a pharmacogenomic algorithm to guide the treatment of depression*[1]. This was the pilot study for next trial in this list: non-randomized, open, single center, comparison of pharmacogenomic guided treatment vs usual treatment for 8 weeks. N=51, 44 analyzed. Dropout rate = 14%. Approximate Delta and SD = 3.5 and 6.0. Approximate delta/SD ratio= 0.58

2) Hall-Flavin DK et al 2013.Utility of integrated pharmacogenomic testing to support the treatment of major depressive disorder in a psychiatric outpatient setting[2]. Non-randomized, open, single center, comparison of pharmacogenomic guided treatment vs usual treatment for 8 weeks. N=227, 165 analyzed. The dropout rate = (27%) with 37% AGT vs 18% TAU dropout. The approximate delta and SD are 3.1 and 7.2. Approximate delta/SD ratio= 0.43

3) Winner JG et al 2013. A prospective, randomized, double-blind study assessing the clinical impact of integrated pharmacogenomic testing for major depressive disorder[3]. Pilot study, randomized, single center, blinded subject and rater, comparison of pharmacogenomic guided treatment vs usual treatment for 10 weeks. (Although described as double-blind, the Winner study does not meet the minimum conventional double-blind standard – blinded investigator and subject) N=51, 49 analyzed. The approximate Week 8 delta/SD ratio=0.47. This ratio was


Confidential Property of Genomind, Inc.

extrapolated from the % SIGH-D-17 improvement results, and is likely an overestimate of the mean SIGH-D-17 improvement ratio.

The three studies are consistent with one another. The small pilot studies show evidence of efficacy, but are too small for reliable inference. Study 1 greatest observed efficacy (Cohen's d =0.58). The second pilot study (Study 3), is more rigorous, adding randomization and rater/subject blinding, with extrapolated Cohen's d = 0.47, not statistically significant.

The largest study (Study 2) shows a somewhat lower Cohen's d (0.43), a more reliable result due to the much larger sample size and the added sensitivity analyses to assess robustness. Accepting that there remains potential bias due to the non-randomized, open design, this study result was used as the best available basis for sample size calculation.

The sample size is calculated for the two-sided alpha=0.05, 90% power test of two independent means of improvement in SIGH-D-17. In this case the sample size depends only on the postulated Cohen's d, and the % dropouts. Note this should be a conservative estimation of sample size, as the actual analysis will use a more efficient MMRM method. However, this conservatism is balanced by the potential underestimation of needed sample size due to design bias in the reference study.

Using the Trial 2 observed difference in SIGH-D-17 improvement and variability, with resulting Cohen's d=0.43, the resulting sample size is 115 per group, with no dropouts. If we assume the dropout rate somewhat less than the reference (23% vs 27% reference), then with the conventional dropout correction, the needed sample size per group is 150, or 300 total randomized subjects. The table below summarizes the design properties of this size study, under varying assumptions for delta, power and % missing Week 8 SIGH-D-17 assessment. The statistical testing significance level is two-sided alpha < 0.05, and the SD is assumed to be the same (7.2).


Confidential Property of Genomind, Inc.

**Detectible* Delta for Varying Power and Missing Data Assumptions**

| Total Sample Size | Subject SD | Delta | Power | % Missing Week 8 |
|---|---|---|---|---|
| 300 | 7.2 | 3.09 | 90% | 23% |
| 300 | 7.2 | 2.67 | 80% | 23% |
| 300 | 7.2 | 2.94 | 90% | 15% |
| 300 | 7.2 | 2.54 | 80% | 15% |
| 300 | 7.2 | 2.85 | 90% | 10% |
| 300 | 7.2 | 2.46 | 80% | 10% |
| 300 | 7.2 | 2.70 | 90% | 0% |
| 300 | 7.2 | 2.34 | 80% | 0% |

*Detectible delta is the smallest postulated "true" difference between means for which 80% (or 90%) of trials would be statistically significant with the given % missing Week 8 data.

All sample size calculations used Stata Version 13 software.

As reiterated above, the reference study is not a definitive foundation for sample size calculation. However, it may be expected the following would hold:

1) The base planning SIGH-D-17 delta, at 3.1, is a high bar for an active vs active treatment depression trial. Commonly, the planning delta for placebo controlled depression trials is 3.0 (SIGH-D-17 units) and active vs active studies can expect delta of 1.5 or less and may often designed as non-inferiority studies. It could be argued that a planning delta should be 2.5 or less.
2) We expect the SD for this study should be less than the reference study. SIGH-D-17 will be assessed by blinded, trained SIGH-D-17 raters, centrally monitored by an independent group. This reduces bias and improves precision, and should result in the primary endpoint being more accurate, with a smaller SD. A reduction of the SD by 15% seems reasonable (SD=6.12).
3) We expect the 8-week dropout rate will be considerably less than the reference, as well as the nominal 23% used in the base sample size calculation. As discussed in Section 8.1, as this trial will institute measures to reduce missing data, consistent with the recommendations by the 2010 NRC Missing Data panel. This, combined with the observed lower dropout rate in active vs active trials, suggests a target of 10% dropouts is realistic.

Assuming the improved rigor of the trial conduct reduces the SD by 15% and the dropout rate to 10%, then with 300 randomized subjects and a 0.05 2-sided test, the “detectable delta” is more realistic: 2.42 with 90% power, and 2.09 with 80% power.

***Sample Size Computations for the Exploratory Elderly MDD Study*:**

*Sample size was not calculated for the Exploratory Elderly MDD Study.*


Confidential Property of Genomind, Inc.

# 5 REFERENCES

1. Hall-Flavin DK, Winner JG et al. Using a pharmacogenomic algorithm to guide the treatment of depression. *Translational psychiatry*. 2012;2(10):e172
2. Hall-Flavin DK, Winner JG et al. Utility of integrated pharmacogenomic testing to support the treatment of major depressive disorder in a psychiatric outpatient setting. Pharmacogenetics and genomics. 2013; 23(10):545-548
3. Winner JG, Carhart JM, et al. A prospective, randomized, double-blind study assessing the clinical impact of integrated pharmacogenomic testing for major depressive disorder. Discovery medicine. 2013;16(89):219-227


Confidential Property of Genomind, Inc.