Genomics in Michigan Impacting Observation or Radiation (G-MINOR)

Protocol CU008
Version: Final 1.6
Date: November 23, 2020

KEY PERSONNEL

Sponsor: **GenomeDx Biosciences Inc.**
[redacted]

Principal Investigators: **Todd Morgan, MD**
Department of Urology
University of Michigan
[redacted]

Michael Cher, MD
Department of Urology
Wayne State University School of Medicine
[redacted]

Data Management: **GenomeDx Biosciences Laboratories**
[redacted]

Statistician(s): **Darby Thompson, PhD**
EMMES Canada
[redacted]

Rodney Dunn, MS
Department of Urology
University of Michigan
[redacted]

Kasra Yousefi, MSc
[redacted]

Chief Medical Officer: Timothy J Triche, MD
Chairman and Chief Medical Officer
GenomeDx Biosciences
[redacted]

Clinical Laboratory: GenomeDx Biosciences Laboratories
[redacted]

Signature of Agreement for Protocol

I agree to conduct this clinical study in accordance with the design and specific provisions of this protocol (CU008, Version Final 1.0) and will only make changes in the protocol after notifying the sponsor.

I will conduct the study in accordance with Good Clinical Practice, the Declaration of Helsinki, and the moral, ethical and scientific principles that justify medical research. The study will be conducted in accordance with all relevant laws and regulations relating to clinical studies and the protection of patients.

Study Principal Investigator:

_____	_____
Signature	Date

DOCUMENT REVISION HISTORY

Version	Date	Replaces	Description of Changes
1.0	2/11/16		
1.1	3/1/16	1.0	Modifications per PRC
1.2	6/15/16	1.1	Modifications per IRB
1.3	11/9/2017	1.2	Enrollment period Modifications per PI
1.4	5/14/2018	1.3	Enrollment period modifications per PI
1.5	4/1/20	1.4	Increased total accrual number per PI
1.6	11/23/20	1.5	Removing the cost related Secondary Objective and Secondary Endpoint; Add new sections, Exploratory Objective and Exploratory Endpoint.

Table of Contents

Signature of Agreement for Protocol 4
Table of Contents 6
1. Introduction and Background 9
2. Study Objectives 10
2.1. Primary Study Objective 10
2.2. Secondary Objectives 10
2.3. Exploratory Objective 11
3. Study Design 11
3.1. Enrollment and Randomization 11
3.2. Genomic Classifier Education Procedures 12
3.3. CAPRA-S and GC Result Reporting 12
3.4. Clinical Data Collection 12
4. Subject Eligibility 14
4.1. Inclusion Criteria: 14
4.2. Exclusion Criteria: 14
5. Endpoints 14
5.1. Primary Endpoint 14
5.2. Secondary Endpoints 15
5.3. Exploratory Endpoint 15
6. Assay Specimen Requirements andLogistics 15
6.1. FFPE Block Selection/Specimen Acquisition 15
6.2. Specimen Preparation and Annotation 15
6.3. Specimen Logistics 16
6.4. Specimen shipping instructions 16
6.5. GenomeDx Decipher® Testing 17
7. Statistical Considerations 17
7.1. Statistical Methods Supporting the Primary Objective 17
7.2. Interim Sample Size Re-Estimation 18
7.3. Statistical Methods Supporting the Secondary Objectives 18
7.4. Power and Sample Size 20
7.5. Simulation settings 20
7.5.1 Power and Sample Size Simulation 20
7.5.2 SimulationResults 21
8. Confidentiality and EthicalIssues 22
9. Scientific Presentation and PublicationPolicy 22

10. Data and Safety Monitoring 22
11. References 23

DEFINITIONS & ACRONYMS

ADT	Androgen Deprivation Therapy
ART	Adjuvant Radiation Therapy
AUC	Area Under Curve
CC	Clinical Classifier
FFPE	Formalin-Fixed, Paraffin-Embedded
GC	Genomic Classifier
GCC	GC augmented with Clinicopathologic
GC-PD	Genomic Classifier Predictive
GCC-PD	GC augmented with Clinicopathologic Predictive
HIPAA	Health Insurance Portability and Accountability Act
MCAR	Missing Completely At Random
PI	Principle Investigator
PSA	Prostate Specific Antigen
RNA	Ribonucleic Acid
RP	Radical Prostatectomy
RT	Radiation Therapy

1. Introduction and Background

Some patients are at high risk of recurrence after radical prostatectomy (RP). Despite being classified as high risk, many patients will never have disease recurrence. Our current tools for predicting who will recur are imprecise, which may explain why the majority of patients at high risk of recurrence do not receive adjuvant therapy despite supporting evidence (Thompson 2009). In a recent study from the SEARCH database, only 67/2670 patients (2.5%) received adjuvant therapy despite 448 (17%) being categorized as high risk (reference). New genomic classifier (GC) assays may help more precisely determine who should be referred for adjuvant therapy after RP.

One such assay is Decipher from GenomeDx Biosciences, a GC assay developed and validated in a post-RP setting as a predictor of eventual metastasis (Erho 2013). Studies have demonstrated independent prognostic value of this GC assay above and beyond standard clinicopathologic parameters reclassifying the clinically high-risk men tested into lower or significantly elevated risk of metastatic disease progression (Karnes 2013, Den 2014, Cooperberg 2015, Klein 2015, Den 2015, Ross 2015, Yamoah 2015). Studies have demonstrated that Decipher low risk patients can safely delay/defer treatment and avoid the adverse side effects of radiation treatment after radical prostatectomy while Decipher high risk patients can benefit from early adjuvant radiation (Den 2014, Den 2015).

In addition to clinical validation, several decision impact studies of Decipher were conducted to investigate a GC's influence on physicians' post-radical prostatectomy treatment recommendations (Badani 2013, Badani 2014, Michalopoulos 2014, Nguyen 2015).

Badani (2013) evaluated the influence on urologist treatment recommendations with GC in comparison to clinicopathologic risk factors alone for patients from a tertiary care center. This prospective, pre-post design study showed that among twenty-one high volume (i.e., mean 180 RP/year) urologic surgeons from 18 US institutions who were asked to make treatment recommendations for 24 prostate cancer patient cases post-RP, 43% changed their treatment recommendation with the addition of genomic information to standard clinicopathologic information.

To further reflect the distribution of test results and risk of metastasis among high-risk patients seen in a broader range of clinical settings other than academic centers, a prospective, multi-center, decision-impact study (ASSESS-D, Badani 2014) evaluated the effect of knowledge of the GC model on urologist treatment decision-making in a population of patients having undergone radical prostatectomy in a community-based practice. One hundred ten patient case histories were available for review by 51 urologists. Overall, 31% (95% CI: 27- 35%) of treatment recommendations changed with knowledge of GC. Treatment intensity was strongly correlated with the GC-predicted probability of metastasis ($p<0.001$) and the GC was the dominant risk factor driving decisions in multivariable analysis (OR=8.6, 95% CI: 5.3-14.3%, $p<0.0001$).

In a prospectively recruited cohort of high-risk post-RP subjects from community-based physicians, Michalopolous (2014) demonstrated that that those classified as GC high-risk were

significantly more likely to be treated (64.3%) than those classified as GC low-risk (4.4%). In addition, the authors demonstrated that the GC score was more influential than clinical variables on adjuvant treatment decisions. In particular, even among men at higher clinical risk (Gleason 4+3 or higher or CAPRA-S ≥ 5), a high GC score resulted in an increase in treatment recommendations. The converse is also seen among clinically low-risk subjects where a low GC score results in a reduction of recommendations for treatment.

These decision impact studies demonstrate the substantial impact a GC has on individual patient management above and beyond clinical variables or nomograms. Further, patient outcomes modeling study showed that Decipher-directed individualized care is associated with a 16% relative increase in the 10-year recurrence-free survival probability (Lobo 2015). However, the clinical utility of GC, and its relation to clinical risk factors, remains unverified in a randomized prospective trial and its ultimate benefit to patients and cost-effectiveness are unclear.

We propose a prospective, randomized trial comparing the frequency of receipt of adjuvant therapy for high-risk RP patients who undergo GC testing vs. those who do not. The study will be performed within the Michigan Urological Surgery Improvement Collaborative (MUSIC), a physician-led quality improvement consortium comprised of the vast majority of urology practices within the state of Michigan

2. Study Objectives

2.1. Primary Study Objective

To determine the impact of GC test results on adjuvant treatment decisions of high-risk post-RP patients with undetectable post-op PSA compared to clinical factors alone. Specifically, determine whether higher GC risk increases the likelihood of receiving adjuvant treatment and lower GC risk decreases the likelihood of receiving adjuvant treatment compared to subjects with similar CAPRA-S score who do not have the GC test.

2.2. Secondary Objectives

1. To determine whether the decision impact of the GC test is of benefit to patients. Specifically, is net benefit of GC use superior to those who do not have the GC test as assessed by decision curve analysis.

2. To determine whether use of the GC test results in decreased likelihood of adjuvant treatment for subjects with high CAPRA-S score (≥5) and increased adjuvant treatment for subjects with low CAPRA-S score (<5). The hypothesis is that there is a non-zero interaction between GC test use and CAPRA-S.

3. To determine whether GC use is associated with earlier use of salvage therapies in CAPRA-S low risk patients and later use of salvage therapy in CAPRA-S high-risk patients.

4. To determine whether use of GC results in improved protection from BCR and metastatic progression.

5. To investigate the impact of GC use on QOL, and whether there is differential QOL impact among CAPRA-S high- and low-risk patients.

6. To determine whether collaborative-wide education and the use of CAPRA-S scores in the control group impacts rates of adjuvant radiotherapy compared to historical controls.

2.3. Exploratory Objective

1. To determine whether it is possible that use of GC will result in an overall reduced net cost to the healthcare system by avoiding treatment in many patients.

3. Study Design

This is a 4-5 year (up to 2 years enrollment , 3-year follow-up) prospective, cluster-crossover, unblinded, study of 550 subjects from at least fourteen centers of the Michigan Urological Surgery Improvement Collaborative (MUSIC), a physician-led quality improvement consortium comprised of the vast majority of urology practices within the state of Michigan.

According to the cluster-crossover design, each participating MUSIC clinical center within MUSIC will be randomized to either a Genomic Classifier care-based (GC) or Usual-Care-based (UC) strategy for a period of 3 months, during which all enrolled subjects within the center will be provided the assigned strategy. Random assignments will be generated centrally by a study statistician and provided to centers immediately before commencing enrollment in each 3-month period. The randomization will be based on a random block size of either 2 or 4. This will ensure that each center will have two periods of GC and two periods of UC during the enrollment period, maintaining study-wide balance and blinding of assignments in subsequent periods.

If enrolled during the GC period, both subjects and their treating physician will be provided GC (from GenomeDx) and CAPRA-S (from MUSIC-Central) to inform their treatment plan. In the UC periods, only the CAPRA-S results, provided by MUSIC-Central, will be available. Blinding of subjects and treating physicians will not be possible given the nature of the information.

Subjects enrolled to the GC arm will receive MUSIC-led education about the GC test. The pros and cons of the GC test need to be described by an unbiased clinician scientist including the past history of the test (published studies).

3.1. Enrollment and Randomization

Due to the prospective nature of this study and the potential for the Decipher test results to materially impact the clinical management of patients participating in the study, all patients must be consented to participate in the study. Eligible patients (Section 4.1) identified by the participating urologist should be consented to discuss the results of the surgical pathology and Decipher results using the Informed Consent Form (ICF).

In the event that the treating urologist decides to terminate a patient's involvement in the study they must notify the sponsor and provide a reason for termination in the electronic case report form.

As described in the preceding section, this is a cluster-randomized trial so individual subjects do not receive a randomized group assignment, but are allocated to the group assigned to the center at the time they enroll.

3.2. Genomic Classifier Education Procedures

It is important to ensure that physician investigators are familiar with interpreting the test results prior to measuring its impact on their treatment recommendations, a set of pre-study regulatory requirements and study training will precede the initiation of this study.

- An overview of the current use of adjuvant and salvage radiotherapy in MUSIC, of the published data on Decipher, and of the study itself will be presented by the PIs at a collaborative-wide meeting. Each participating practice must be present at this meeting or review this content directly with the PIs. The presentation will include a balanced discussion of the current evidence and interpretation of test results in order to avoid potential bias.
- Each physician must review educational materials provided by GenomeDx Biosciences on the Decipher test
- Patient education materials surrounding the study will be provided by MUSIC to eachof the practices for patients to review.

3.3. CAPRA-S and GC Result Reporting

The results of the GenomeDx Decipher® test will be communicated in the form of a test report that will be provided to the ordering physician by FAX and a hardcopy is shipped with UPS. The MUSIC coordinating center will generate a CAPRA-S report for all participating patients that will be provided by FAX or e-mail to each practice. The treating urologist will review the CAPRA-S score with the patient.

3.4. Clinical Data Collection

Where available, participating physicians in the study will be asked to provide the following clinicopathologic data in the provided case report form (CRF) for subjects meeting the inclusion criteria below. All data submitted to GenomeDx will utilize a case ID number and will be de-identified by the site prior to submission.

1. Ethnicity
2. Age at radical prostatectomy
3. Pre-operative PSA level: last pre-operative PSA level before radicalprostatectomy
4. Lymph node status
5. Clinical and pathological stage
6. Clinical and pathological Gleason score (primary, secondary, and tertiary pattern if available)

7. Surgical margin status
8. Extra-prostatic extension
9. Seminal vesicle invasion
10. All available post-op PSA levels
11. Recommended treatment or management option for the patient after surgery (detailed by type of treatment recommended)
12. Dates of actual treatments
13. Patient outcome

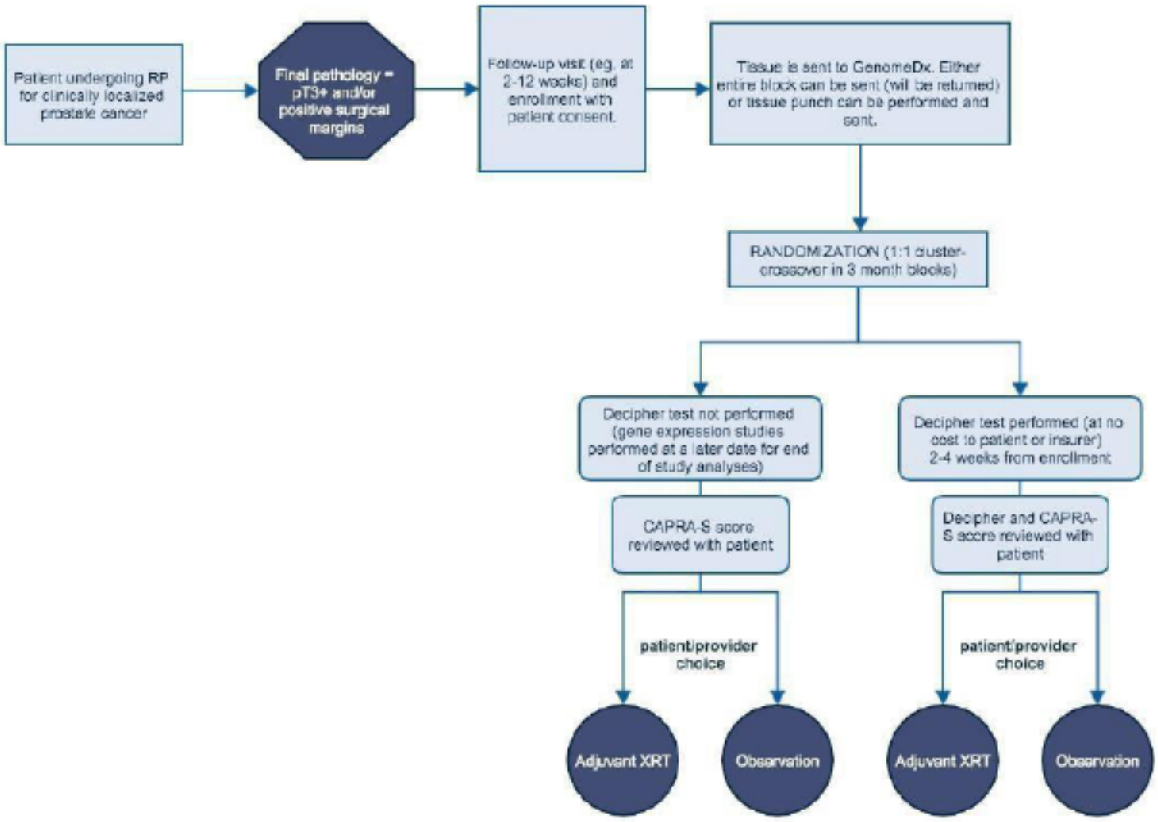

Figure 1: Study flow diagram

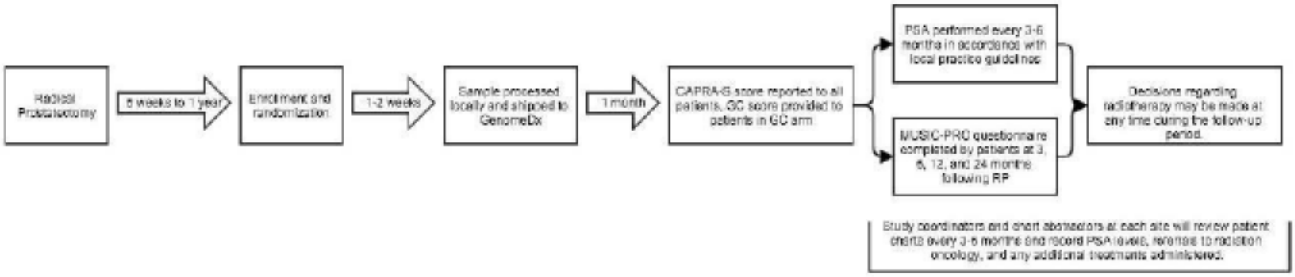

Figure 2: Patient flow timeline

4. Subject Eligibility

4.1. Inclusion Criteria:

Individuals satisfying all of the following criteria will be eligible to participate.

1. Prostate cancer patients who have undergone radical prostatectomy(RP)
2. PSA <0.1 ng/ml at enrollment
3. At least one of the following:
    a. pT3 (SVI or EPE), ***or***
    b. Positive Surgical Margins
4. Tumor tissue (FFPE block) available for processing
5. RP within 1 year of enrollment
6. 18 years of age or older

4.2. Exclusion Criteria:

Individuals who satisfy ***any*** of the following criteria will not be eligible to participate.

1. Have regional or distant metastatic disease
2. Received any radiation or hormone therapy (neo-adjuvant, adjuvant, or salvage)
3. Node positive
4. Subjects who do not have banked specimens

5. Endpoints

5.1. Primary Endpoint

The Primary Endpoint will be whether the subject receives any adjuvant therapy (radiation and/or hormone therapy). Adjuvant will be defined as preceding BCR (i.e.: PSA ≥0.2 ng/ml) and within 18 months of RP. Patients must have a PSA within 6 months prior to the receipt of radiotherapy.

5.2. Secondary Endpoints

The following secondary endpoints will be utilized:

1. Time (from randomization) to adjuvant treatment administration
2. Time (from randomization) to salvage treatment administration
3. Biochemical Recurrence (BCR) at 3 years
4. Metastatic disease (regional or distant) at 3 years
5. QOL measures currently collected as part of the MUSIC-PRO initiative (http://musicurology.com/pro/) at 3, 6, 12, and 24 months after surgery

5.3. Exploratory Endpoint

Cost comparison/difference between GC and UC care at 3 years after surgery

6. Assay Specimen Requirements and Logistics

6.1. FFPE Block Selection/Specimen Acquisition

FFPE (formalin fixed paraffin embedded) tissues from RP prostate tumor sample blocks will be collected from patients in both the GC and UC arms for transcriptome-wide expression analysis.

The FFPE block from the RP that exhibits a) highest Gleason Score, AND b) highest percent tumor involvement will be selected for sampling with at least 60% tumor by surface area. The pathologist will mark the H&E slide that will be used to guide punching. The marked area should preferentially sample tumor with the highest Gleason grade (e.g., in a 4+3 lesion, the pattern 4 component should be enriched in the punch specimen). Two (2) X 1.5 mm in diameter punches will be sampled from the RP FFPE block (e.g., using a tissue microarray core device or a disposable punch tool, to be provided by GenomeDx). Each punch will be placed in a separate microfuge tube and stored at 4°C until shipping.

6.2. Specimen Preparation and Annotation

GenomeDx will ship specimen kits for each patient sample. Pathologists will have the option of sending the FFPE block or using the provided punch tool to perform two core biopsy punches per specimen to be placed in the provided microfuge tube or slides in the provided slide carriers.

If FFPE block is sent, place the representative block into small zip lock bag provided.

1. Select one FFPE block representative of the tumor.

Choose a block with at least 0.5 cm2 of tumor tissue with the highest Gleason grade of tumor. If multiple blocks exist, select the block with the highest percent tumor cells by area.

2. Please do not submit an H&E slide. GenomeDx will prepare an H&E slide on site.

If punch tool core biopsy method for FFPE radical prostatectomy blocks is used:

1. Identify a H&E slide that closely matches the block face. Cut and H&E stain a new slide, if multiple sections have already been cut from the block.

2. Circle the area with the highest Gleason grade on the corresponding H&E slide. a) The punch area should have a minimum of 60%tumor (by surface area)

3. Using the punch tool provided in the specimen kit, take two punches from the FFPE block which corresponds to the area circled on the H&E slide.

a) Two (2) x 1.5 mm diameter cores

b) Minimum length of tissue in the punched cores should be 2mm

4. Place each cores in a separate microcentrifuge tubes. Close the caps and seal tubes with Parafilm.

5. Document that punch contains > 60% tumor content on Specimen ProfileForm.

6. Affix provided barcodes only to tubes containing specimens and number each tubeto correspond with the Specimen Profile Form.

Duplicate cores may be requested by GenomeDx if specimen provided does not yield either sufficient RNA, cDNA or fails microarray quality control.

Microfuge tubes to be labeled with unique de-identified patient and specimen numbers (assigned by GenomeDx and matching the case numbers assigned along with the clinicopathologic details).

If submitting slides:

Please send five (5) x 5um thick unstained, uncharged sections plus one (1) H&E stained slide from the block as defined in our block selection criteria.

6.3. Specimen Logistics

1. Place FFPE block or microfuge tubes containing specimens into the foam insertin the specimen kit box provided
2. Place a copy of the completed requisition form in the plastic bag provided.
3. Place the provided coldpack on top of the foam insert, close the box and seal in a clear bag.

6.4. Specimen shipping instructions

Specimen kits are to be sent in a shipping box via FedEx (FedEx Account # to be provided upon

site enrollment) to the following address:

GenomeDx Biosciences

[redacted]

6.5. GenomeDx Decipher® Testing

GenomeDx will extract and purify RNA from submitted tumor specimens. If the sample fails to yield sufficient RNA (at least 100 ng) GenomeDx will request a duplicate specimen from the submitting physician (preferably from a different FFPE block for that patient). GenomeDx will then generate cDNA, amplify, label and hybridize it to oligonucleotide arrays. The Decipher gene signature tested will remain unchanged throughout the course of the study, such that all patients will undergo the identical assay.

GenomeDx will then perform QC on the raw microarray data. Samples that fail array QC will be re-amplified (provided there is sufficient RNA remaining). If there is insufficient RNA or the 2nd amplification fails, the duplicate sample will be extracted. After generation of the array data and QC analysis, GenomeDx will return (on dry ice) RNA or specimen cores not used for amplification to the pathologist or discard appropriately.

7. Statistical Considerations

7.1. Statistical Methods Supporting the Primary Objective

The Primary Objective of this study is to determine the impact of GC test result on adjuvant treatment decisions compared to clinical factors alone. This will be evaluated by testing for an effect of GC score in a regression model which adjusts for CAPRA-S risk score. As a cluster-crossover trial, a generalized linear mixed effects model will be used to model the probability of a subject receiving adjuvant treatment within center, in a specific period of the study. Since GC score is a subject-level covariate, an individual-level regression model is utilized rather than a cluster-level aggregated analysis (Campbell 2014, Chapter 6). Specifically, the model to be fit is

$$logit\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right)=\beta_0+\beta_1 G_i+\beta_2\left(G_i * GC_i\right)+\beta_3 \text{CAPRA-S}_i+\gamma_j+\phi_k$$

where π_{ijk} is the probability of treatment for the i^{th} subject enrolled in the j^{th} cluster (ie: center) during the k^{th} period, G_i is the group to which the subject is enrolled (GC=1, UC=0), CAPRA-S_i is that subjects' CAPRA-S score (0-12), GC_i is the GC score (0-1), $\gamma_j \sim N(0,\sigma_\gamma^2)$ and $\phi_k \sim N(0,\sigma_\phi^2)$ are random cluster and period effects, respectively. The null hypothesis $H_0: \beta_2 = 0$ will be tested using a Type III test (SAS PROC GLIMMIX) at a 0.05 significance level. 95% confidence intervals using robust variance estimates will also be calculated.

In addition to the model described above, an additional model including a group-by-period interaction term will be fit to ensure that group assignment in a period is not effected by preceding or subsequent periods. For the purposes of this analysis, the period effect will be

considered fixed. This may detect carry-over effects if behaviors or attitudes of physicians within sites persist or change over the course of the study or if changes in patient education materials or their presentation change from one period to the next. Should this interaction term be significant, an alternative analysis utilizing non-consecutive periods may be used.

Descriptive statistics for various quantities, including raw treatment rates (and 95% confidence intervals) of adjuvant treatment and type of treatment (hormone and radiation) for discrete risk categories for both CAPRA-S (low-risk 0-4, high-risk ≥5) and GC (low-risk ≤ 0.45, high-risk > 0.45) as well as by clinical center and period. Model-based predicted probabilities of treatment will be calculated for various combinations of CAPRA-S and GC high and low risk along with associated confidence intervals.

7.2. Interim Sample Size Re-Estimation

A confirmation of assumptions underlying the model described in Section 7.1 and the Power and Sample Size calculations in Section 7.4 will be conducted after no fewer than 150 subjects have been enrolled and adjuvant treatment determinations have been made, several parameters will be estimated based on the accumulated data, but no hypothesis testing will be performed and Type I error will be nominally effected (Lake 2002). The following parameters will be estimated and compared to those assumed in Section 7.4, and the impact on the estimated sample size determined through simulation:

- average enrollment by center, including variance of enrollment and enrollment by period
- intra-cluster correlation (ICC)
- overall rate of treatment

Based on the estimated parameters, the number of centers or sample size may be adjusted while maintaining the requisite 80% power. If simulations indicate that a sample size increase of >20% will be required to provide adequate power to achieve the Primary Objective, the study may be terminated due to futility.

7.3. Statistical Methods Supporting the Secondary Objectives

All tests will be two sided at a 0.05 significance level. Individuals who are lost to follow-up will be considered censored in time to event analyses. Treatment plans available at the time of drop-out will be considered the final treatment plans.

Secondary Objective 1: To determine whether the net benefit to subjects in the GC arm is superior to the UC arm. This will be determined by directly calculating the net benefit (Vickers and Elkin 2006) across a range of treatment thresholds, Decision Curve Analysis (DCA). DCA is a method for assessing prognostic models that incorporates clinical consequences and the potential probability thresholds at which the likelihood of a given outcome (eg: progression to metastatic disease) may influence the treatment decision of a patient and their physician. For example, one individual may choose treatment if there is a 5% chance of a given outcome while another individual may not be influenced unless the likelihood of that outcome is as high as 10 or 15%. Thus, this methodology allows for how different patients may differentially weigh false negative or false positive results. The calculated net benefit is plotted against threshold

probabilities, and different prediction tools (eg. CAPRA-S and GC) can be compared across these probabilities. If GC is not superior to UC across all thresholds, the region of superiority will be evaluated to determine if it contains a clinically meaningful ratio of harm due to unnecessary treatment versus missed treatment.

Secondary Objective 2: To determine whether use of the GC test result results in decreased likelihood of treatment for CAPRA-S high and increased likelihood of treatment for CAPRA-S low, a modeling approach similar to that of supporting the Primary Objective will be used, but with a GC-by-CAPRA-S interaction term as follows:

$$logit\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right)=\beta_0+\beta_1 G_i+\beta_2\left(G_i * GC_i\right)+\beta_3 \text{CAPRA-S}_i+\beta_4\left(G_i * \text{CAPRA}_i * GC_i\right)+\gamma_j+\phi_k$$

where parameters are defined as previously. The test of $H_0:\beta_4=0$ will be performed using a Type III test (SAS PROC GLIMMIX). The direction of the interaction will be evaluated graphically with the probability of treatment based on the model and for empirical probabilities using pre-defined GC and CAPRA-S cutoffs.

Secondary Objective 3: Time-to salvage therapy initiation will be modeled using a Cox formulation of a Frailty regression model with a similar formulation (ie: main and random effects) to that used for the primary analysis. Kaplan-Meier curves for GC and UC arms will be plotted, as well as for combinations of high- and low-risk GC and CAPRA-S groups.

Secondary Objective 4: Time-to BCR and metastatic progression will be evaluated similarly to the model described for Secondary Objective 3.

Secondary Objective 5: Summary statistics for all QOL instruments and corresponding confidence intervals will be calculated and compared for the GC and UC subjects and shown graphically over time. Validated instruments are used to measure erectile function, urinary function, sexual interest, and sexual satisfaction. Scores for each domain will be scaled with a minimum to maximum range of 0-100. Comparisons between GC and UC subjects will be made at each timepoint (3, 6, 12 and 24 months after surgery) and on each domain and overall score (VAS 0-10), using an ANCOVA analysis including adjustment for time since surgery. Should the assumptions for ANCOVA not be met (strongly skewed data or substantial ceiling or floor effect) a semiparametric ANCOVA model will be used (as in Buja 1989) to adjust for time-from surgery. If the distribution of time-from surgery is similar in the two groups, a Mann-Whitney test for differences will be used to corroborate the ANCOVA model. For this secondary objective, there will be no adjustment for multiple comparisons across domains or time points.

Exploratory Objective: Using a variety of budget models the cost of care for patients in the GC vs. UC arms will be compared out to 3 years from surgery. This will include estimated costs for adjuvant therapy. Differences in total cost, and differential cost per patient (UC vs GC) will be calculated and compared using either standard t-tests if data is normally distributed or appropriate non-parametric test if data is right skewed.

7.4. Power and Sample Size

There are no existing computer packages that can determine the power or sample size for a parameter in a mixed-effect logistic model from a cluster-crossover trial. As a result, power and sample size estimates were determined through large-scale simulation utilizing the model and analysis method described in Section 7.1.

7.5. Simulation settings

In order to determine plausible values of the parameters of the model described in Section 7.1, several clinical utility studies of the GC were pooled and utilized in a re-sampling routine to determine likely values of parameters for the current study. Although none of the previous clinical utility studies were randomized comparative trials, the DECIDE (Badani 2013), ASSESS-D (Badani 2014), and PRO-ACT Michalopolous (2014) studies did involve real high-risk patient cases (only those that were pT3 or SM+ are included here) from community and tertiary care facilities and can be expected to provide similar clinical and genomic characteristics to those eligible for this trial. In all studies, treatment recommendations were provided by physicians initially without GC risk scores then (in a blinded fashion) with knowledge of both clinical and GC risk.

There are a total of 530 recommendations with and without GC available from these clinical utility studies. To approximate a parallel-group (GC and UC) randomized trial, for each run of the re-sampling routine, the data was randomly halved, taking half of the recommendations made with knowledge of the GC (n = 265) and the other half with recommendations made without knowledge of the GC (n = 265). Ten-thousand such data sets were generated and to each, a generalized linear model of the probability of treatment was fit, similar to that described in Section 7.1 but involving only the fixed effects. Across these 10,000 samples, the mean ORs of GC score, CAPRA-S score, and arm (GC vs UC) are, respectively, 1.47, 1.15, and 0.24. The overall rate of adjuvant treatment of CAPRA-S high-risk patients in MUSIC clinics is ~8%.

7.5.1 Power and Sample Size Simulation

Generating CAPRA-S and GC Scores

Simulation was used to determine the sample size necessary to have adequate power to reject the null hypothesis, $H_0 : \beta_2 = 0$, based on the model in Section 7.1. CAPRA-S and GC risk scores for the simulated datasets were based on the distributions observed in the PRO-ACT clinical utility study (Michalopolous 2014). The distribution of CAPRA-S risk in PRO-ACT clinical is similar to that of MUSIC (62.7% vs 55.8% low risk and 44.2% vs 37.3% high risk). The distribution of GC risk scores among MUSIC patients is unknown, but due to the similarity in CAPRA-S scores, it is assumed that it will also be somewhat similar to PRO-ACT.

To ensure a) the GC score distribution is comparable to PRO-ACT, b) the CAPRA-S risk distribution is as similar as possible to MUSIC, c) the true variability of the continuous GC score, and d) the underlying correlation between GC and CAPRA-S risk is maintained, the following procedure was used to generate GC and CAPRA-S scores for each simulated individual:

1. A random percentile is generated (0-100%) and rounded to the nearest5%.
2. PRO-ACT scores with GC score (and corresponding CAPRA-S score) within that percentile range (eg: 20-25th percentile) are then selected.
3. A GC score is then generated as a uniform a random number between the maximum and minimum of the selected PRO-ACT GC scores.
4. An initial CAPRA-S score is then randomly sampled from the selected CAPRA-S scores. The initial CAPRA-S score is then incremented by 1-point (to a maximum of 12) with 75% probability to up-adjust the slightly lower risk in PRO-ACT.

Generating Enrollments

Based on data available from MUSIC-central, the peak performing clinical centers exceeded 10 radical prostatectomy cases per month, with case-loads below 5 per month for the 4th and 5^{h} centers. Enrollments per site*month were generated using a Poisson process using the following conservative means: Site 1: 10, Site 2: 10, Site 3: 5, Site 4: 5, Site 6-8: 4, Site 9 and higher: 2.

Sites were randomized to enroll subjects to the GC or UC arm for each period using a fixed-block size enrollment (2, 3, or 4 depending on the randomization ratio of the simulation). Period length varied from month to year.

Intra-Class Correlations

In other clinical utility studies on GC (eg: Badani 2014) physicians reviewed several cases. With physicians as clusters, the ICC in this study is approximately 38% using methods of Wu (2012). In somewhat different scenarios, and with alternative methods of calculation Adams (2004) identified the IQR of ICCs as 3.5-28.4% in cluster-randomized trials. Assuming that the ICC for an entire clinical center will be less than for a single physician, and given the findings of Adams, an ICC of 23% was selected for the simulation – given the other parameter settings of the model, specified above, this is equivalent to setting $\sigma^2_T = 1$. A small intra-period (across sites) is also incorporated (ICC = 3%) to include the possibility of a temporal correlation across sites - potentially related to changes in educational materials or seasonal trends.

7.5.2 SimulationResults

Partial simulation results are listed in Table 1. Five-thousand simulated datasets were generated and analyzed at each setting – simulation error is ~ 0.5%. Various combinations of the site-ICC, number of sites, duration of randomization periods, and sample size were attempted. A study of 350 subjects at 14 clusters (centers) enrolled over 3-month periods equally randomized to GC or UC, will have 83% power to detect a significant CAPRA-S-adjusted effect of GC score on the likelihood of adjuvant therapy at the 0.05 level. Such a study will require 8-12 months (ie: 3-4 periods) for recruitment.

Minor perturbations in the ICC (15-35%), number of sites (5-10), and the randomization period duration (1-4 months) did not effect the power by more than 1%, with a sample size of 350. However, including 4 or fewer sites will result in a power < 80% and while inclusion of at 8 or

more sites will maintain the desired power (> 80%), inclusion of these lower-recruiting sites will not appreciably decrease the enrollment period.

Additional components influencing the required sample size are the total event rate, average and variance of site-period recruitment, and actual approximation of the randomization ratio (will not be exactly 1:1). These, as well as the site-ICC, will be evaluated as described in Section 7.2, Sample Size Re-Estimation.

Table 1: Power comparisons for various sample sizes.

Site-ICC	Number of Sites	Period Duration (months)	Approx. Duration of Enrollment (months)	Sample Size	Power
23%	8	3	6	200	47%
23%	8	3	6	250	60%
23%	8	3	7	300	77%
23%	8	3	8	350	83%
23%	8	3	9	400	88%
23%	8	3	12	500	93%

8. Confidentiality and Ethical Issues

The Principal Investigators are responsible for conducting this Study Protocol. The study will be reviewed and approved by an IRB/EC and appropriate Scientific Review Committees prior to initiation.

All participating physicians must sign a confidentiality and non-disclosure agreement.

9. Scientific Presentation and Publication Policy

Presentation and publication of the results will be based upon appropriate analysis and review of the complete data, and the usual rules for determining authorship will guide scientific presentation and publication of the results of this study. Written approval from GenomeDx Biosciences is required prior to disclosing any confidential information relative to this study.

10. Data and Safety Monitoring

This study will be monitored in accordance with the NCI approved University of Michigan Comprehensive Cancer Center Data and Safety Monitoring Plan.

The study team will meet every six months or more frequently depending on the activity of the protocol. The discussion will include matters related to the safety of study participants (SAE/UaP reporting), validity and integrity of the data, enrollment rate relative to expectations, characteristics of participants, retention of participants, adherence to the protocol (potential or real protocol deviations) and data completeness. At these regular meetings, the protocol specific Data and Safety Monitoring Report form will be completed and signed by the Principal Investigator or by one of the co-investigators.

Data and Safety Monitoring Reports will be submitted to the University of Michigan Comprehensive Cancer Center Data and Safety Monitoring Committee (DSMC) every six months for independent review.

11. References

Adams, Geoffrey, Martin C. Gulliford, Obioha C. Ukoumunne, Sandra Eldridge, Susan Chinn, and Michael J. Campbell. "Patterns of intra-cluster correlation from primary care research to inform study design and analysis." Journal of clinical epidemiology 57, no. 8 (2004): 785-794.

Badani, Ketan, Darby JS Thompson, Christine Buerki, Elai Davicioni, Jill Garrison, Mercedeh Ghadessi, Anirban P. Mitra, Penelope J. Wood, and John Hornberger. "Impact of a genomic

classifier of metastatic risk on postoperative treatment recommendations for prostate cancer patients: a report from the DECIDE study group." Oncotarget 2013; 4(4): 600-9.

Buja, A., Hastie, T. J. and Tibshirani, R. J. (1989). Linear smoothers and additive models (with discussion). Annals of Statistics, 17, 453–555.

Badani, Ketan K., Darby J. Thompson, Gordon Brown, Daniel Holmes, Naveen Kella, David Albala, Amar Singh, Christine Buerki, Elai Davicioni, and John Hornberger. "Effect of a genomic classifier test on clinical practice decisions for patients with high-risk prostate cancer after surgery." British Journal of Urology Intl 2014; 115(3): 419-429.

Campbell, Michael J., and Stephen J. Walters. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. John Wiley & Sons, 2014.

Erho, Nicholas, Anamaria Crisan, Ismael A. Vergara, Anirban P. Mitra, Mercedeh Ghadessi, Christine Buerki, Eric J. Bergstralh et al. "Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy." PLOS ONE 2013; 8(6):e66855.

Heo, Moonseong, and Andrew C. Leon. "Comparison of statistical methods for analysis of clustered binary observations." Statistics in medicine 24.6 (2005): 911-923.

Karnes, R. Jeffrey, Eric J. Bergstralh, Elai Davicioni, Mercedeh Ghadessi, Christine Buerki, Anirban P. Mitra, Anamaria Crisan et al. "Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population." The Journal of Urology 2013; 190:2047-2053.

Den, R. et al., "Genomic Prostate Cancer Classifier Predicts Biochemical Failure and Metastases in Patients After Postoperative Radiation Therapy." Int J Radiat Oncol Biol Phys. 2014; 89(5):1038-46.

Cooperberg, M.R., Davicioni, E., Crisan, A., Jenkins, R.B., Ghadessi, M., and Karnes, R.J., Combined Value of Validated Clinical and Genomic Risk Stratification Tools for Predicting Prostate Cancer Mortality in a High-risk Prostatectomy Cohort. European Urology, 2015 Feb; 67(2):326-333.

Klein, Eric A., Kasra Yousefi, Zaid Haddad, Voleak Choeurng, Christine Buerki, Andrew J. Stephenson, Jianbo Li, Michael W. Kattan, Cristina Magi-Galluzzi, and Elai Davicioni. "A Genomic Classifier Improves Prediction of Metastatic Disease Within 5 Years After Surgery in Node-negative High-risk Prostate Cancer Patients Managed by Radical Prostatectomy Without Adjuvant Therapy."

European Urology 2015; 67(4): 778-786.

Den, R.B., Yousefi, K., Trabulsi, E.J., Abdollah, F., Choeurng, V., Feng, F.Y., Dicker, A.P., Lallas, C.D., Gomella, L.G., Davcioni, E., and Karnes, R.J., A genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. Journal of Clinical Oncology 2015; 33:944-951.Ross AE, Johnson, M.H., Yousefi, K., Davicioni, E., Netto, G.J., Marchionni, L., Fedor, H.L., Glavaris, S., Choeurng, V., Buerki, C., Erho, N., Lam. L.L., Humphreys, E.B., Faraj, S., Bezerra, S.M., Han, M., Partin, A.W., Trock, B.J., Schaeffer, E.M., Tissue Based Genomics Augment Post-Prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High- Risk Men European Urology, 2015; DOI:10.1016/j.eururo.2015.05.042.

Yamoah, K., Johnson, M.H., Choeurng, V., Faisal, F.A., Yousefi, K., Haddad, Z., Ross, A.E., Alshalafa, M., Den, R., Lal, P., Feldman, M., Dicker, A.P., Klein, E.A., Davicioni, E.,Rebbeck, T.R., Schaeffer, E.M., A novel biomarker signature which may predict aggressive disease in African-American men with prostate cancer. Journal of Clinical Oncology, 2015; DOI: 10.1200/JCO.2014.59.8912.

Lake, Stephen, Erin Kammann, Neil Klar, and Rebecca Betensky. "Sample size re-estimation in cluster randomization trials." Statistics in medicine 21, no. 10 (2002): 1337-1350.

Michalopoulos, Steven N., et al. "Influence of a genomic classifier on postoperative treatment decisions in high-risk prostate cancer patients: Results from the PRO-ACT study." Curr Med Res Opin 2014; 30(8):1547-56.

Nguyen, Paul L., H. Shin, K. Yousefi, D. J. Thompson, J. Hornberger, A. S. Hyatt, K. K. Badani, T. Morgan, and F. Y. Feng. "Impact of a Genomic Classifier of Metastatic Risk on Post prostatectomy Treatment Recommendations by Radiation Oncologists and Urologists." Urology 2015; 86(1): 35-40.

Lobo, J.M., Dicker, A.P., Buerki, C., Davicioni, E., Karnes, R.J., Jenkins, R.B., Patel, N., Den, R.B., and Showalter, T.N., Evaluating clinical impact of a genomic classifier in prostate cancer using individualized decision analysis. PLoS One. 2015 Apr 2;10(3):e0116866. doi: 10.1371/journal.pone.0116866. eCollection 2015.

Thompson, Ian M., Catherine M. Tangen, Jorge Paradelo, M. Scott Lucia, Gary Miller, Dean Troyer, Edward Messing et al. "Adjuvant radiotherapy for pathological T3N0M0 prostate cancer significantly reduces risk of metastases and improves survival: long-term followup of a randomized clinical trial." The Journal of urology 181, no. 3 (2009): 956-962. DOI:10.1016/j.juro.2008.11.032

Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making. 2006 Nov-Dec;26(6):565-74.

Wu, Sheng, Catherine M. Crespi, and Weng Kee Wong. "Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials." Contemporary clinical trials 33.5 (2012): 869-880.