**Electrophysiological Characterization of Subclinical and Overt Hypertrophic Cardiomyopathy by Magnetic Resonance Imaging-Guided Electrocardiography**

**Supplemental Statistical Methods:** p1

**Supplemental Results:** p1-3

**Supplemental Table 1:** p4

**Supplemental Table 2:** p5

**Supplemental Table 3:** p6

**Supplemental Table 4:** p7

**Supplemental Figure 1:** p8

**References:** p8

**SUPPLEMENTAL STATISTICAL METHODS:**

**Multivariable Regression Models**

The normality of model residuals was validated using visual inspection of histogram and Q-Q plots. Multiple linear regression modelling involving fractionation required log-transformation, however other ECGI variables showed no need for transformation. Variance inflation factors ≤3 excluded significant multicollinearity. Homoscedasticity and independence of residuals were verified through scatterplots and the Durbin-Watson statistic, respectively, both yielding satisfactory results.

**Machine Learning Methods**

Optimal cutoff values to discriminate between subclinical HCM and healthy volunteers for each ECGI parameter separately were initially obtained using R package ‘cutpointr’ maximize metric function. Next, supervised machine learning (ML) with the support vector machine (SVM) classification was used to build an ECGI biomarker panel for HCM as previously described (1) given that the basic assumption of non-multicollinearity between the 12-ECGI parameters for a logistic regression could not be met. SVMs are a set of effective, supervised non-parametric ML techniques that recognize patterns in data, and are especially suited to two-group separation challenges (2)(3). Another advantage of SVMs, is their relative resistance to over-fitting given that they use regularization (4). The goal of our exploratory SVM model was to combine all the available ECGI features to predict which phenotypic category (subclinical HCM or control) a participant belonged to. For each model, the 4 available kernels in package ‘e1071’ were tested and individually tuned using pre-set cost (0.1,0.2,0.4,0.6,0.8,1,10,100,1000) and gamma (0.1,0.2,0.4,0.6,0.8,1,2,3,4) arguments to select the best model (**Supplemental Table 5**). The final SVM was trained using the subclinical HCM vs control cohort with a polynomial kernel tuned to cost 10 and gamma 0.1 followed by a 10-fold cross-validation to correct for overoptimism. SVM performance was summarized in terms of its confusion matrix of classifications and area under the receiver operating characteristics curve (AUC, ROC) calculated using package ‘ROCR’. 95% Confidence intervals for AUCs were computed with 2000 stratified bootstrap replicates using ‘pROC’. A similar SVM and ROC approach explored whether the ECGI electrotype in patients with overt HCM could discriminate low risk from high/intermediate risk patients (from the ESC Risk score), including those with prior documented NSVT. The latter SVM used a radial kernel tuned to cost 1 and gamma 0.6. Variable importance in the final models was determined using package ‘rminer’. **(Supplemental Table 7 & 8)**.

The R script is available at <https://github.com/georgejoyucl/HCM-ECGI>.

**SUPPLEMENTAL RESULTS**

**G+LVH+ and matched HV** (Supplemental Table 1)

To compare ECGI changes in G+LVH+ to health, an older age-, sex- and ethnicity-matched healthy volunteer group were studied (n=23). Compared to older HV, G+LVH+ were similar in age (G+LVH+ vs HV respectively: 52(37-59) vs 44(37-55) years, *p*=0.48), sex (35 men (69%) vs 15(65%), *p*=0.77) and ethnicity (43 White (84%) vs 19(83%), *p*=0.85). G+LVH+ had similar AT but greater AT dispersion, more fractionation, prolonged ventricular repolarization (elevated ARIc and RTc) and more spatially heterogenous repolarization (elevated GRTc) (**Supplementall Table 1**).

**ECGI relationships in pooled HV (including older controls, n=37)**

Age was associated with mean AT (*r*s=0.37 [95% confidence interval: 0.04­, 0.6] *p*=0.025) and inversely associated with mean GAT (rs=-0.37 [-0.05, -0.6] *p*=0.023). Age inversely associated with repolarization duration (ARIc: *r*s=-0.32 [-0.009,-0.6] *p*=0.05; RTc: *r*s=-0.35 [-0.01, -0.6] *p*=0.039) and inversely associated with GRTc (*r*s=-0.39 [-0.07, -0.6] *p*=0.017).

Male sex was associated with mean AT (*r*s=0.35 [0.02, 0.6] *p*=0.033), and inversely with repolarization duration (ARIc: *r*s=-0.44 [-0.1, -0.7] *p*=0.007; RTc: *r*s=-0.50 [-0.2, -0.7] *p*=0.002). Male sex was inversely associated with mean GRTc (*r*s=-0.38 [ -0.05, -0.6] *p*=0.022).

Ethnicity, BSA and BMI were not associated with any ECGI parameter.

**Overt HCM with NSVT subgroup analysis**

16(15%) of patients with overt HCM had NSVT. Compared to those without NSVT, these patients were similar in terms of age (*p*=0.07), sex (*p*=0.46) and ethnicity (*p*=0.46), but they had higher MWT (19.0(17-24)mm vs 13.5(10-18)mm *p*=0.01) and LGE volume (16.6(9-32)g vs 1.6(0-13)g, *p*<0.001) and more spatially heterogenous conduction (GATmax: *p*=0.012). Differences remained after adjustment for MWT and LGE volume (ß=0.026 [0.1, 0.5] *p*=0.016).

**Gene Specific Analysis**

There were no differences in ECGI parameters between carriers of *MYH7* vs *MYBPC3* variants in both subclinical and overt disease. However, across all genotype positive participants (LVH+/LVH-), *MYBPC3* variant carriers had less fractionation than non-*MYBPC3* variants (*p*=0.007). Differences persisted after adjusting for age, sex, LGE volume and MWT (*p*=0.003). *MYBPC3* carriers also had significantly lower RTc (*p*=0.02) and GATmax (*p*=0.046) compared to non-*MYBPC3* variant carriers on univariate analysis, but these were attenuated in fully adjusted models.

**Drugs and ECGI parameters**

68% (71) of participants with overt HCM were on at least one medication (including non-cardiac). Compared to those without medication, patients with overt HCM on medication were older (58[50-60] vs 48[32-48] years *p*<0.001) but were similar in terms of sex (*p*=0.81), BMI (*p*=0.13), ethnicity, MWT (*p*=0.09), LGE burden (*p*=0.71) and presence of LVOTO (*p*=0.46).

In order to explore whether overt HCM had ECGI changes without medication usage, we compared this group (*n*=33) to age-, sex- and ethnicity-matched healthy volunteers (*n*=22). Compared to matched HVs, overt HCM patients without medication were of similar age (overt HCM vs HV respectively: 48[32-58] vs 41[36-53] years *p*=0.34), sex distribution (21% female [7] vs 32% [7] *p*=0.38) and ethnicity (85% White [28] vs 82%[18] *p*=0.77).

Compared to healthy volunteers, drug-free overt HCM had prolonged repolarization (longer ARIc and RTc) and greater maximal repolarization gradients. All other ECGI parameters were similar (**Supplemental Table 2**).

We performed additional analysis using univariate and multivariable regression in overt HCM with MWT, LGE, age, sex and drug class as covariates. Drug classes were beta-blockers, non-dihydropyridine calcium channel blockers, dihydropyridine calcium channel blockers, disopyramide and amiodarone. We found no associations between ECGI parameters and the differing drug classes.

**LV Morphology and ECGI relationships**

Compared to isolated basal LVH, reverse septal morphology had higher signal amplitude and more prolonged repolarization. Compared to Other morphology (defined as concentric or mixed apical LVH), reverse septal curvature had lower signal amplitude. Other morphology had higher signal amplitude, more prolonged activation and repolarization times compared to isolated basal septal LVH. **(Supplemental Table 3)**

**Machine Learning Classification of ECGI HCM subtypes**

As expected, diagnostic cut-points for individual ECGI biomarkers tasked with discriminating subclinical HCM from controls were not sufficiently accurate, with the exception of max GRTc and mean AT with accuracies of 73% [63–82%] and 79% [69–86%] respectively (**Supplemental Table 4**). Therefore, we explored the use of supervised machine learning applied to the combined 12-biomarker ECGI electrotype in subclinical HCM and HV. This SVM differentiated subclinical HCM from HV with an AUC of 0.96 (bootstrap 95% confidence interval: 94–98%; sensitivity 100% [93–100%], specificity 91% [76–98%], positive predictive value [PPV] 96% [87–99%], negative predictive value [NPV] 100% [87–100%], balanced accuracy 95.7%, **Supplemental Table 6** and **Supplemental Figure 1**) and an accuracy of 80% after 10-fold cross-validation [73–85%]. This ECGI biomarker panel was able to identify HCM patients at greater risk of SCD (because of prior NSVT or intermediate/high ESC SCD risk status) compared to low-risk patients, with an AUC of 0.97 (bootstrap 95% CI: 96–98%; sensitivity 94% [71–99%], specificity 100% [95–100%], PPV 100% [77–100%], NPV [99–93%], balanced accuracy 97.2%, **Supplemental Table 6**) and an accuracy of 82% after 10-fold cross-validation [78–86%].

**Supplemental Table 1 – ECGI biomarkers in G+LVH+ vs matched healthy volunteers.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  |
|  |  | **HV (n=23)** | **G+LVH+ (n=51)** | ***p* value** |
| Amp, mV | | 1.7(1.1-1.8) | 1.6(1.2-1.9) | 0.96 |
| Fractionation (n/1000) | | 0(0-21) | 13(2-27) | **0.003** |
| Mean AT, ms | | 39(34-43) | 41(37-45) | 0.28 |
| ΔAT, ms | | 136(61-188) | 183(162-208) | **<0.001** |
| Mean RTc, ms | | 272(261-281) | 322(296-340) | **<0.001** |
| ΔRTc, ms | | 150(119-170) | 165(150-178) | **0.023** |
| Mean ARIc, ms | | 232±21 | 276±29 | **<0.001** |
| ΔARIc, ms | | 161(141-194) | 183(162-208) | 0.053 |
| mean GAT, ms/mm | | 0.39(0.25-0.46) | 0.40(0.35-0.46) | 0.07 |
| max GAT, ms/mm | | 4.7(3.9-5.9) | 5.1(4.2-5.8) | 0.37 |
| mean GRTc, ms/mm | | 0.9±0.3 | 1.0±0.3 | **0.019** |
| max GRTc, ms/mm | | 9.0(7.9-10.5) | 11.3(9.9-12.7) | **<0.001** |

Values are median(interquartile range) or mean±standard deviation.

*Amp, amplitude; AT, activation time; Δ, dispersion; G+LVH+, gene variant positive left ventricular hypertrophy positive; HV, healthy volunteer; max, maximum; RTc, repolarization time corrected for heart rate; ARIc, activation recovery interval corrected; G, gradient.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supplemental Table 2 –** **Drug-Free HCM vs matched healthy volunteers.** | | | |  |
|  |  | **Healthy Volunteers (n=22)** | **Drug-Free Overt HCM (n=33)** | ***p* value** |
| Amp, mV | | 1.6(1.1-1.8) | 1.6(1.2-1.9) | 0.57 |
| Fractionation (n/1000) | | 15(1-30) | 7(1-25) | 0.22 |
| Mean AT, ms | | 40(33-43) | 41(36-44) | 0.36 |
| ΔAT, ms | | 157(62-193) | 180(152-194) | 0.08 |
| Mean RTc, ms | | 267(257-288) | 317(298-343) | **<0.001** |
| ΔRTc, ms | | 153(126-174) | 161(143-185) | 0.31 |
| Mean ARIc, ms | | 232±22 | 275±27 | **<0.001** |
| ΔARIc, ms | | 171(153-203) | 179(156-196) | 0.88 |
| mean GAT, ms/mm | | 0.39(0.26-0.46) | 0.36(0.30-0.41) | 0.96 |
| max GAT, ms/mm | | 4.6(3.9-6.0) | 4.7(4.2-5.3) | 0.75 |
| mean GRTc, ms/mm | | 1.0(0.7-1.2) | 1.0(0.8-1.2) | 0.46 |
| max GRTc, ms/mm | | 9.1(8.3-10.6) | 10.9(9.5-13.7) | **0.008** |

Values are median(interquartile range) or mean±standard deviation.

*Abbreviations as in* ***Supplemental Table 1****.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | ***p*-values** |  |
|  | **Isolated Basal Septal LVH**  **(n=52)** | **Reverse Septal Curvature**  **(n=37)** | **Other (Apical/Concentric)**  **(n=15)** | **ANOVA** | **Isolated basal LVH vs reverse septal curvature** | **Isolated basal LVH vs Other** | **Reverse septal curvature vs Other** |
| Amp, mV | 1.5(1.2-1.9) | 1.8(1.4-2.4) | 2.0 (1.7-2. 6) | **<0.001** | **0.009** | **<0.001** | **0.021** |
| Fractionation (n/1000) | 5(0-22) | 1(0-36) | 1(0-25) | 0.77 |  |  |  |
| Mean AT, ms | 41(36-45) | 41(38-45) | 49(38-54) | **0.002** | 0.203 | **0.008** | 0.11 |
| ΔAT, ms | 181(156-197) | 196(162-213) | 188(167-214) | 0.38 |  |  |  |
| Mean RTc, ms | 314(296-332) | 330(314-347) | 341(326-347) | **<0.001** | **0.009** | **<0.001** | 0.25 |
| ΔRTc, ms | 159(145-184) | 171(157-194) | 171(163-180) | 0.21 |  |  |  |
| Mean ARIc, ms | 183±31 | 199±40 | 188±29 | **0.013** | **0.019** | **0.014** | 0.57 |
| ΔARIc, ms | 183(158-202) | 197(173-225) | 188(162-211) | 0.093 |  |  |  |
| mean GAT, ms/mm | 0.40(0.30-0.49) | 0.42(0.35-0.53) | 0.40(0.32-0.56) | 0.84 |  |  |  |
| max GAT, ms/mm | 5.1(4.3-5.8) | 5.0(4.3-6.0) | 5.0(4.2-6.4) | 0.96 |  |  |  |
| mean GRTc, ms/mm | 1.01(0.86-1.19) | 1.08(0.77-1.25) | 1.01(0.82-1.24) | 0.89 |  |  |  |
| max GRTc, ms/mm | 10.9(9.3-13.3) | 11.5(9.6-13.5) | 11.2(9.5-13.0) | 0.9 |  |  |  |

**Supplemental Table 3 – Relationship between LV morphology and ECGI parameters.**

Values are median(interquartile range) or mean±standard deviation.

*Abbreviations as in* ***Supplemental Table 1****, Other is mixed ASH-apical LVH or concentric LVH*

**Supplemental Table 4 – Diagnostic cut-points for individual ECGI biomarkers.**

|  |  |  |
| --- | --- | --- |
|  | **Diagnostic ECGI cut-off points** | |
| **Conduction** |  |  |
| Mean AT, ms |  | |
|  |  |  |
| **Repolarization** |  |  |
| Max GRTc, ms/mm |  | |

Accuracy defined as = True Positive + True Negative / True Positive + True Negative + False Positive + False Negative.

**Abbreviations as in Supplemental Table 1.**

**Supplemental Table 5 – Comparative kernel performance after best-model tune.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Kernel** | **Optimum Cost** | **Optimum Gamma** | **AUC** |
| **Subclinical Model** | **Linear** | 0.4 | 0.1 | 0.79 |
| **Radial basis** | 1 | 0.1 | 0.80 |
| **Polynomial** | 10 | 0.1 | **0.96** |
| **Sigmoid** | 1 | 0.1 | 0.69 |
| **HCM-Risk Model** | **Linear** | 0.1 | 0.1 | 0.50 |
| **Radial basis** | 1 | 0.6 | **0.97** |
| **Polynomial** | 0.1 | 0.1 | 0.78 |
| **Sigmoid** | 0.6 | 0.6 | 0.42 |

AUC highlighted in bold indicate the final kernel and model parameters selected.

*AUC, area under the curve.*

**Supplemental Table 6 – Confusion matrices for the two final SVM models.**

|  |  |  |  |
| --- | --- | --- | --- |
| **SVM model distinguishing subclinical HCM from controls** | | | |
|  | **Control** | **Subclinical HCM** | **Totals** |
| **Test Positive** | 3 | 68 | 71 |
| **Test Negative** | 32 | 0 | 32 |
| **Totals** | 35 | 68 | 103 |
| **SVM model distinguishing NSVT+ or intermediate/high-risk from low-risk patients** | | | |
|  | **Low-risk** | **Intermediate/high-risk** | **Totals** |
| **Test Positive** | 0 | 17 | 17 |
| **Test Negative** | 85 | 1 | 86 |
| **Totals** | 85 | 18 | 103 |

**Supplemental Table 7 – Contribution of each ECGI determinant to the predictive capacity of the subclinical vs control SVM model.**

|  |  |
| --- | --- |
| **Ranked ECGI Parameter** | **Importance coefficient** |
| Mean AT | 0.306 |
| Mean GRTc | 0.263 |
| Fractionation | 0.167 |
| Max GRTc | 0.067 |
| ΔAT | 0.062 |
| Max GAT | 0.062 |
| Mean RTc | 0.038 |
| ΔARIc | 0.017 |
| ΔRTc | 0.011 |
| Mean Amplitude | 0.005 |
| Mean GAT | 0.002 |
| Mean ARIc | 0.001 |

**Supplemental Table 8 – Contribution of each ECGI determinant to the predictive capacity of the low vs intermediate/high-risk HCM SVM model.**

|  |  |
| --- | --- |
| **Ranked ECGI Parameter** | **Importance coefficient** |
| Max GAT | 0.197 |
| ΔAT | 0.155 |
| Mean AT | 0.127 |
| Mean ARIc | 0.086 |
| Mean GRTc | 0.077 |
| Fractionation | 0.076 |
| ΔARIc | 0.074 |
| Mean GAT | 0.073 |
| Max GRTc | 0.049 |
| Mean Amplitude | 0.035 |
| ΔRTc | 0.026 |
| Mean RTc | 0.025 |

**Supplemental Figure 1. ECGI signature potentially distinguishes subclinical HCM from controls and identifies patients at greater risk of SCD.**

ROC curve (**a**) showing the performance of the support vector machine supervised machine learning method applied to all combined ECGI biomarkers in distinguishing patients with subclinical HCM from healthy volunteers (2 subclinical HCM patients and 2 HV were excluded due to missing ECGI parameters). ROC curve (**b**) showing the ability of this ECGI electrotype to identify patients with HCM at intermediate/high risk for SCD or with prior documented NSVT from low-risk patients (1 overt HCM patient excluded due to missing ECGI parameters), with an accuracy of 99% (AUC 0.97, bootstrap 95% CI: 96–98%), which on 10-fold cross validation maintained a mean accuracy across the 10 folds of 82% [78–86%].



**REFERENCES**

1. Hollander Z, Dai DLY, Putko BN, Yogasundaram H, Wilson-McManus JE, Thompson RB, et al. Gender-specific plasma proteomic biomarkers in patients with Anderson–Fabry disease. Eur J Heart Fail. 2015 Mar 1;17(3):291–300.

2. Liu HX, Zhang RS, Luan F, Yao XJ, Liu MC, Hu ZD, et al. Diagnosing Breast Cancer Based on Support Vector Machines. J Chem Inf Comput Sci. 2003 May 1;43(3):900–7.

3. Swan AL, Mobasheri A, Allaway D, Liddell S, Bacardit J. Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. OMICS J Integr Biol. 2013 Dec 1;17(12):595–610.

4. Cawley G, Talbot N. Preventing Over-Fitting during Model Selection via Bayesian Regularisation of the Hyper-Parameters. J Mach Learn Res. 2007 Apr 1;8:841–61.