**补充材料**

# 电子病历和基因组学网络:过去、现在和未来

海伦娜Kuivaniemi暗利戈特斯曼,医学博士,医学博士,博士,杰拉德跺脚,博士,w . Andrew Faucett女士,Rongling,医学博士,博士,泰瑞Manolio,医学博士,博士Saskia c·桑德森博士,医学博士约瑟夫·Kannry Randi Zinberg,女士,公司治理文化,梅丽莎·a . Basford MBA,穆雷聪明,博士,博士,大卫·j·凯里雷克斯·l·奇泽姆博士,克里斯托弗·g .槽医学博士认为,医学博士约翰·j·康诺利博士David Crosslin约书亚·c·丹尼,医学博士,卡洛斯·j·加利西亚人,医学博士乔纳森·l·海恩斯博士,医学博士,博士,约翰·哈雷Hakon hankonarson博士说医学博士,Gail P. Jarvik，医学博士，Isaac Kohane，医学博士，Iftikhar J. Kullo，医学博士，Eric B. Larson，医学硕士，Catherine B. McCarty，医学博士，Marylyn D. Ritchie，医学博士，Dan Roden，医学博士，Maureen E. Smith, MS, Erwin P. Bottinger，医学博士，Marc S. Williams，医学博士和eMERGE Network

# 紧急情况ii地点的说明

图1显示了9个研究小组的位置、它们的附属站点、一个协调中心以及组成emerger - ii的服务和支持中心。eMERGE网络的活动轮廓如图2所示，网络的组织结构如图3所示。除以下具体信息外，生物传感器、EMR系统和基因分型项目的详细情况如表1所示。表S1列出了每个位点的主要目的，表S2列出了每个位点所选择的主要和次要表型。

## 辛辛那提儿童医院医疗中心(CCHMC)和波士顿儿童医院(BCH)

CCHMC和BCH共同努力成为一个单一的新兴站点。它们的生物库解决方案各不相同。

CCHMC现在在门诊和住院病人注册中获得“选择加入”的许可，允许在研究中使用那些本来会被丢弃的生物材料。目前大约有15000个冷冻组织和15000个DNA样本被保存。DNA样本正以每月超过2000份的速度累积。BCH正在招募1万名儿童和家庭成员加入基因合作计划，收集样本，并广泛同意返回结果，包括一个基于web的界面，向患者及其监护人展示结果。BCH的这个收集将参与许多关于结果返回的研究。

机构合并为一个站点，为BCH和CCHMC提供了协调和解决机构间问题的机会，从而使算法扩展到其他新兴机构将更加简单。第一个目标是构建信息基础设施，使一个机构开发的算法能够迅速应用于另一个机构。

这个CCHMC/BCH的eMERGE II站点建立在五年的合作历史上，特别是在与病人emr相关的信息学方面，这是大部分的eMERGE II的基础。所有CCHMC/BCH站点的GWAS和EMR数据都链接到共享数据仓库平台i2b2,1-4，并可在共享的全基因组关联数据库GWADB中进行分析。GWAS数据来源于各种来源，包括临床服务

(细胞遗传学)，正常对照，研究者发起的表型特异性研究，和BCH的研究连接-基因伙伴关系(TGP)，一个关注于了解复杂疾病的根本原因的倡议。正在开发的是i2b2中定义表现型的对象的EMR视图，该视图使用SHRINE5分布式查询系统的修改。这将使跨机构的实时病人选择在全国范围内进行。

儿童早期严重肥胖(2到5岁)被选为表现型认为早发性疾病更可能是不那么复杂,因此比成人更基因齐次形式的严重肥胖的分析需要大量的情况下,许多基因的发现与温和的影响。肥胖流行病也在影响着这些儿童。在中国疾病预防控制中心，有3000多名儿童似乎符合纳入的初步标准，并且按照世卫组织的标准，他们的BMI名列前0.1%。

自闭症正在以惊人的频率折磨着美国儿童。CCHMC和BCH都有大型诊所和有天赋的临床医生，试图控制和改善这些儿童及其家庭所面临的严重问题。这是一个复杂的表现型算法的发展。复杂性被用来应用启发式和机器学习的方法都被用来探索这种情况的遗传。

对于次要表型，CCHMC/BCH网站正准备将算法应用于哮喘和2型糖尿病，这两种疾病在儿童中都是严重的问题。

1. Kohane是，Churchill SE, Murphy SN。国家规模的翻译引擎:整合生物学和临床的信息学。J Am Med Inform Assoc. 2012;19:181-185。

2. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R，等。为基因组时代的发现研究服务的医疗保健企业。基因组研究》2009;19:1675 - 1681。

3. 小羽。利用电子健康记录来推动疾病基因组学的发现。Nat Rev Genet. 2011;12:417-428。

4. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V，等。来自电子健康记录的多种族队列中自身抗体阳性和阴性类风湿关节炎风险的遗传基础。Am J Hum Genet. 2011;88:57-69。

5. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S等。共享卫生研究信息网络(SHRINE):用于临床数据存储库的联邦查询工具原型。J Am Med 2009;16:624-630

## 费城儿童医院(CHOP)

费城儿童医院(CHOP)的应用基因组学中心(CAG)已经建立了世界上最大的儿童生物基因库之一，并正在积极研究50多种影响从出生到21岁的儿童的疾病。招聘——由10名全职护士和抽血医生维持——主要以人口为基础，尽管一些针对疾病的工作同时进行。CAG biorepository目前总共有> 70000个内部样本，包括> 5000个完整的trios。我们通过协作项目(包括几千个trio)维护另外100,000个样本。生物样品库正以每年12- 15000个样品的速度增长。所有样本/数据在接触点都经过加密和条形码处理。超过85%的参与者同意/同意重新联系以参与进一步的研究。所有(内部)参与者均同意/同意回顾性和前瞻性的纵向电子病历。电子病历由EpicCare®(自2001年起以印章形式实施)处理，存储库参与者的电子病历平均长度为>5.5年。emr由一系列内部项目提供的数据丰富，其中包括一项后续研究，该研究对10,000名生物或医学参与者进行了3小时的神经行为测试，其中1,500人完成了神经成像研究。此外，所有的参与者都要完成一份包括自我报告的人口、健康和家庭历史数据的登记问卷。样品由来自REMP (Oberdiessbach，瑞士)的一个基于机器人的系统存储和管理，该系统促进了快速的数据共享，并极大地降低了放置/调用错误的可能性。超过99%的样品可以通过CLIA/CAP认证或下一代测序进行再基因分型。该中心的分析能力旨在识别医学疾病谱系中的遗传变异，我们在>50疾病领域有积极的项目。生物样品库以其非洲裔美国人样本的比例(43%)而著称。

切出现项目的主要目标是1)建立现有举措出现表型相关算法适应儿科军团(脂质在这方面主要表型),2)来定义新的表型相关性成人和儿科军团(哮喘、注意缺陷多动障碍、过敏性皮肤炎,和胃食管返流疾病主要表型在这方面),3)进行GWAS，将EMR数据共享对患者隐私的风险降到最低;4)制定进行研究的同意和社区咨询程序;5)开始将基因组研究结果纳入临床护理。

## Geisinger卫生系统

盖辛格卫生系统(GHS)是一个综合的、全面的卫生保健提供系统，主要服务于宾夕法尼亚州北部、中部和东北部的大量、稳定的农村人口。Geisinger拥有一个功能齐全的综合EMR系统，在EMR和健康信息技术的应用方面是公认的领导者。为了利用卫生系统的资产进行基因组医学研究，Geisinger于2006年启动了一项生物数据库计划，即MyCode®项目，该项目是患者样本(血液、DNA、血清和组织)的中央存储库，可与EMR链接，以保护患者信息机密性的方式广泛应用于研究

Geisinger eGenomic Medicine (GeM)项目的目标是:1)识别与腹主动脉瘤(AAA)、2种极端肥胖、3种及相关疾病相关的遗传变异;2)利用基于emr的临床决策支持工具，开发将基因组数据纳入临床护理的方法;3)确认居住在农村地区的患者对基因检测结果的社会文化关注，并提供患者和医生教育。将遗传数据集成到EMR的策略是基于Geisinger的临床转化团队开发的流程进行建模的，该团队是一个专门的部门，致力于在患者护理中创建支持it的改进。采用“护理缺口”战略，对临床领域进行评估，以确定可纳入EMR的IT流程，从而改善患者护理。Geisinger的ELSI咨询委员会将与eMERGE网络和NHGRI一起制定政策和指导方针，将结果反馈给参与者。它还将为参与者和临床医生创建教育材料。

1. Gerhard GS, Carey DJ, Steele GD, Jr.基因组医学电子健康记录。摘自:Willard HF, Ginsburg GS, eds。基因组和个性化医疗。卷1。2、华豪出版社:学术出版社;2013:287 - 294。

2. Gretarsdottir S, Baas AF, Thorleifsson G等。全基因组关联研究确定了DAB2IP基因中的一个序列变异，该变异可导致腹主动脉瘤的易感性。Nat麝猫。2010;42:692 - 697。

3. Wood GC, Chu X, Manney C, Strodel W, Petrick A, Gabrielsen J, Seiler J, Carey D, Argyropoulos G, Benotti P, Still CD, Gerhard GS。支持电子健康记录的肥胖数据库。BMC Med通知Decis Mak 2012;12:45

康合作社登记的患者，该集团健康合作社是一个总部位于华盛顿西雅图的大型、区域性、集成的交付系统

服务超过62万人。GH emergence biobank包括年龄在> 50岁的受试者来自GH的两个正在进行的生物反应器:1)阿尔茨海默病患者登记/成人思维变化(ADPR/ACT)研究;和2)西北遗传医学研究所(NWIGM)的前瞻性生物存储。现有的eMERGE II队列包括来自这些生物仓库的5299名受试者，其中3561人已经进行了基因分型，他们的数据构成了eMERGE-II项目当前的分析样本。另有1738名受试者有外显子组芯片，但没有GWAS数据。所有受试者均有长期GH EMR。

西雅图，与其他网络站点，将继续进行GWAS分析多种新的表型鉴定通过EMR。该网站计划探索的表型包括与感染性疾病易感性(BuGWAS)相关的表型，特别是与艰难梭菌易感性、水痘带状疱疹病毒的再活化、真菌甲感染(onchomychosis)相关的表型。此外，我们正在研究GWAS鉴定的染色体异常与骨髓疾病的关系。最后，西雅图的研究人员将生成和研究药物遗传数据的实施。通过与医疗系统领导者的一对一访谈，与医生和患者的焦点小组，以及原型开发和测试，西雅图的调查人员正在评估将基因组信息整合到临床护理的挑战。GH、华盛顿大学和弗雷德·哈金森癌症研究中心将通过知识、技术和最佳实践的交流，在产生基于电子病历的GWAS和与临床护理相结合的基础上进行合作和扩展。我们与阿尔茨海默病遗传学协会有特定的合作，并与外部遗传和基因组研究。

马什菲尔德诊所基金会、艾森提亚农村卫生研究所和宾夕法尼亚州立大学

马什菲尔德/艾森提亚/宾夕法尼亚州立大学为他们的eMERGE II项目提出了三个目标:1)开发和验证电子算法眼科疾病和治疗眼科疾病的功效和实现其他表型算法在出现网络发达,2)利用GWAS数据可供近6000名研究对象年龄在50岁及以上在和额外的20 - 25000例出现网络进行基因的发现为眼科条件和眼科药物基因学,3)与一般社区和临床医生开展咨询活动，以将GWAS结果纳入电子健康记录，为卫生保健决策提供信息。关于主要表型的信息可以在表S2中找到。除了使用SNP数据进行GWAS分析外，CNV数据也正在进行清理，将用于出血量1(白内障，高密度脂蛋白)和出血量2的结果。焦点小组讨论采用了与急诊i1相同的方法，对象包括在Marshfield诊所有初级保健医生的普通社区成员、PMRP生物样本库的参与者、2和Marshfield诊所初级保健医生。讨论围绕着基因研究结果的返回以及为患者和提供者提供的支持信息。PMRP社区咨询小组每年召开几次会议，就涌现II的各个方面提供咨询意见，而PMRP小组每年通过两份简报了解研究活动

1. 麦卡锡CA，加伯A, Reeser JC, Fost NC。研究通讯、社区和伦理咨询委员会以及焦点小组讨论为大型生物银行提供持续的反馈。医学杂志2011;155:737-741。

2. 马什菲尔德诊所个性化医学研究项目(PMRP):设计、方法和大型生物样本库的招募。个性化医学2005;2:49 - 79。

3. 麦卡蒂·卡，查普曼·斯通·D，詹皮特罗·PF, Fost NC, PMRP社区咨询小组。基于人群的DNA生物库的社区咨询和交流:马歇尔菲尔德诊所个性化医学研究项目。Am J Med Genet 2008;3026-3033。

## 梅奥诊所

梅奥的主要目标是开展基于emr的多种心血管表型的GWAS1研究，并在社区的持续指导下开发相关的基因组研究结果转化为临床实践的方法。提出的Mayo eMERGE-II队列(n=6,916)包括3,769例eMERGE-I患者和另外3,147人，其中大多数(90%)是Illumina 660W-Quad平台基因分型。大部分Mayo患者(>95%)具有欧洲血统。梅奥调查人员在完成的过程中EMR-based GWAS识别常见的遗传变异的影响:a) inter-individual心肺适能的变化,一个强大的标记的不良结果,和b)对两种常见重大公共卫生重要性的条件:静脉血栓栓塞和心脏衰竭。针对这些表型的验证和可运输的表型分型算法正在开发中，并通过一个开源库(PheKB)向科学界广泛提供。梅奥临床试验II期的第二个主要目标是进行随机临床试验，以调查患者对基因型知情冠心病(CHD)风险的反应。研究人员将从社区中招募150名有10-y冠心病风险的患者，对他们进行clia认证的多种冠心病易感性snp基因分型，并使用专门设计的风险沟通工具将结果和对冠心病风险的影响传达给患者和他们的提供者。沟通的有效性以及患者对风险的理解、对信息的反应和计划中的生活方式的改变将在与冠心病风险沟通后不久和之后的几个时间点通过访谈和调查进行评估。与此同时，梅奥团队正在研究信息学方法，将基因组信息和相关的临床决策支持工具纳入EMR。

1. Kullo IJ, Fan J, Pathak J, Savova GK, Ali Z，降落伞CG。利用信息学进行遗传研究:利用电子医疗记录使外周动脉疾病的全基因组关联研究成为可能。《医学通报》2010年第17卷第568-574页。(PMCID: 2995686)

2. Kullo IJ, Jarvik GP, Manolio TA, Williams MS, Roden DM.利用电子健康记录实现基因组医学。医学版:23018749

## 西奈山医学院

西奈山医院和医学院组成了西奈山医疗中心，这是一个大型的学术医疗中心，每年接待近6万名住院病人和60多万名门诊病人。西奈山的病人主要居住在曼哈顿北部的城市社区，那里的社会经济、种族和宗教都是多样化的。2007年，西奈山建立了查尔斯·布朗夫曼个性化医学研究所(IPM)，并于2007年9月启动了BioMe™生物样本库项目，旨在促进数据驱动和基于基因的医疗个性化。BioMe™是一项正在进行的、前瞻性的、以医院为基础的人群研究，已登记了来自西奈山临床护理设施的23,000多名自愿参与者。1 .参与者同意a)使用其完整的EMR进行表型和暴露数据，b)未来再次联系以参与额外的IRB批准的研究，c)共享去识别的表型和基因型数据。BioMe™组群代表了西奈山当地社区的种族和种族多样性，有25%的非洲血统(AA)， 30%的欧洲血统(EA)， 36%的西班牙裔或拉丁裔血统(HL)， 9%的其他血统。2011年，西奈山以“多元社区基因组医学生物存储”研究项目加入了eMERGE II网络。到2012年12月，GWA基因分型(OmniExpress+Exome芯片)已经扩展到15000多名BioMe™参与者。西奈山已投入大量资源开发由IPM学院领导的复杂信息管理系统，使医疗点结合基因型信息，并在西奈山的EpicCare®EHR提供临床决策支持。

西奈山涌现二项目的目标和目标集中在以下领域:

基因组医学的社区参与教育和研究;表型验证表型算法在慢性肾脏疾病(CKD)及其进展(CKD进展)和药物性肝损伤(DILI)中的发展和传播;通过提供大量遗传多样性患者样本，支持新兴市场II GWAS项目;与西奈山基因组信息EHR整合系统的开发和实施分享经验;包括eMERGE II网络临床护理、药物遗传学支持和APOL1基因组风险信息在非洲裔患者高血压和肾脏护理管理中的本地实施在内的基因组医学试点示范项目。

1. 太波，太美，唐丽，等人。来自曼哈顿一所大学医学中心的病人的遗传背景:对个性化医疗的启示。《公共科学图书馆•综合》。2011;6:e19166。

## 西北大学

西北大学(NU)的出血点是建立在NUgene生物信息库的基础上的，通过该信息库，自愿参与者的DNA样本与来自住院部和门诊护理机构的临床信息相联系，并补充来自参与者问卷的数据。参与者同意将他们的编码DNA样本和数据分发给第三方研究人员进行的范围广泛的遗传研究。

实际上，我们所有的住院和门诊数据都是通过电子方式获取的。门诊记录保存在EpiCare®EMR中，住院记录保存在Cerner Powerchart®1中

在eMERGE I中，NU领导了II型糖尿病、2型哮喘、3类脂质、4类和身高的表现型算法的开发，并实现了其他网站开发的算法。超过4950名NUgene参与者被作为eMERGE I的一部分和NICHD多囊卵巢综合征研究的对照进行了基因分型。NU领导了对非洲裔美国人和欧洲裔美国人T2D GWAS的分析。所有T2D浮现样本的超级分析成功复制了TCF7L2，这是迄今发现的最一致的T2D风险等位基因。此外，对非裔美国人样本的分析还发现了一些有趣的潜在联系(尚未发表)。NU还领导了对非裔美国人低密度脂蛋白胆固醇的分析，发现APOE 4有很强的保护变异。NU公司采用了一种分三个阶段的混合方法来咨询不同社区的遗传研究，并共享GWAS和表型数据，其中包括NUgene biobank的参与者、公众和IRBs。5、6 NU是为计划进行高通量基因组研究和共享数据的生物库开发模型同意语言的主要新兴站点，并领导撰写了一篇关于社区参与努力的比较论文

eMERGE II项目建立在NU的广泛经验之上，包括在研究和开发临床决策支持工具方面使用EHR，以及通过eMERGE I活动将基因组数据整合到临床护理中的基因型数据。9、10本项目的具体目标包括:1)拓展电子表型算法的开发与实现，检测新的基因组关联;2)向医生和患者咨询临床相关基因型的实际应用情况，并告知临床活动;3)开发将遗传变异数据集成到我们的商用EHR的技术方法;4)评估关键转化因素的影响，如监管障碍、EHR决策支持工具，以及医生和患者关于基因组信息的教育，并传播经验教训和最佳实践建议。

1. McCarty CA, Chisholm RL，降落伞CG等人。

2. Kho AN, Hayes MG, Rasmussen-Torvik L，等。在全基因组关联研究中，使用不同的电子医疗记录系统来识别2型糖尿病的遗传风险。J Am Med Inform Ass 2012;19:212-8

3. 药物标签上的药物基因组生物标记。(见2012年7月13日，http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacgenetic s/ucm083378.htm)例如，自然主义探究。贝弗利山:圣人;1985.

4. eMERGE网络:一个连接到电子病历数据的生物存储库联盟，用于进行BMC Med Genom 2011的基因组研究;

5. 遗传研究的广泛数据共享:机构审查委员会专家的观点。IRB 2011; 33:1-5。

6. Lemke AA, Wolf WA, hebert -贝鲁特J, Smith ME。公众和生物库参与者对基因研究参与和数据共享的态度。公共卫生基因组学2010;13:368-77。

7. 电子病历和基因组学(eMERGE)网络同意和社区咨询工作组知情同意工作组:模型同意语言。2009.

http://www.genome.gov/Pages/PolicyEthics/InformedConsent/eMERGEModelLanguage2009-12-15.pdf

8. Lemke AA, Wu JT, Waudby C, Pulley J, Somkin CP, Trinidad SB.社区参与生物银行:来自新兴网络的经验。基因组学Soc政策2010;6:35-52。

9. Kho AN, Pacheco JA, Peissig PL等。用于遗传研究的电子医疗记录:eMERGE联盟的结果。Sci Transl Med 2011;3:79re1。

10. Persell SD, Kaiser D, Dolan NC等。实施基于电子健康记录的质量改进系统后性能的变化。医疗保健2011;49:117-25。

## 范德比尔特大学

Vanderbilt为他们的eMERGE II项目提出了四个目标:1)通过选择和构建表现型算法的原则方法来加速表现型提取算法的发展;2)研究高预测疾病和疾病结局的基因型组合;3)研究患者的基因检测视角，指导药物处方;4)开发工具，最大限度地共享基因组信息，同时保护隐私。可能改变治疗的药物遗传表型尤其值得关注，表S2列出了主要表型。发现的一个重要部分是理解与遗传变异相关的表型范围。因此，他们正在更广泛地部署基于emr的全表型关联研究(PheWAS)1，以广泛地研究已发现的变异的多效性，并指导进一步的表型研究。他们的eMERGE II项目建立在PREDICT(药物基因组资源，用于在护理和治疗中增强决策)项目2的基础上，该项目旨在前瞻性地测试有风险接受药物治疗的患者，这些药物具有已知的药物遗传影响。该项目目前为氯吡格雷、华法林和辛伐他汀提供基于Illumina VeraCode ADME平台测试的计算机化决策支持咨询。截至2012年12月，已有超过1万名患者接受了预测测试。他们已经表明，接触已知药物遗传影响的药物是一个常见的事件，65%的正常门诊人口接受至少一种药物与食品和药物管理局断言的遗传影响

1. Denny, J. C.等人。PheWAS:论证了全表型扫描发现基因疾病关联的可行性。生物信息学26:1205 2010;1210年。

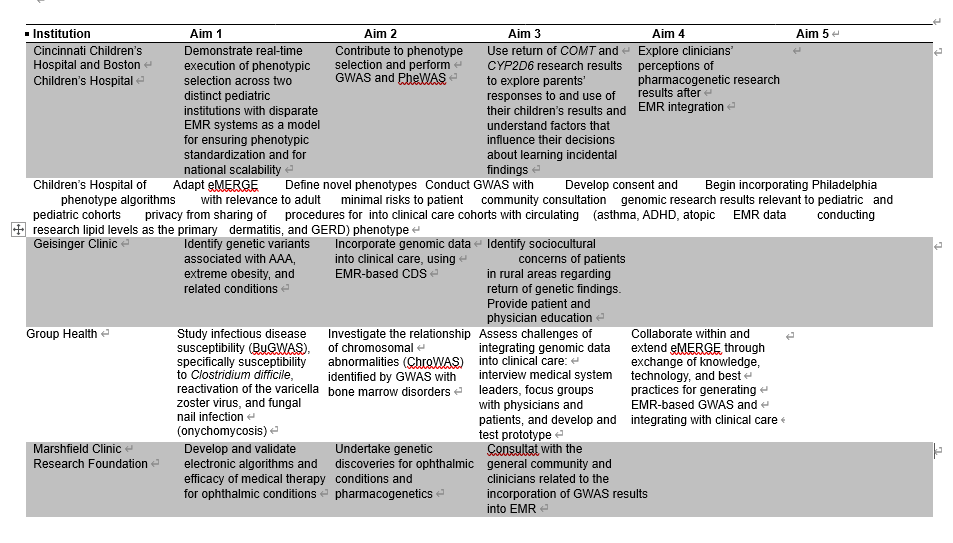
2. 滑轮，j.m.等。前瞻性基因分型在个性化医疗中的应用:范德比尔特预测项目的设计。中国新药杂志Therap。doi: 10.1038 / clpt.2011.371

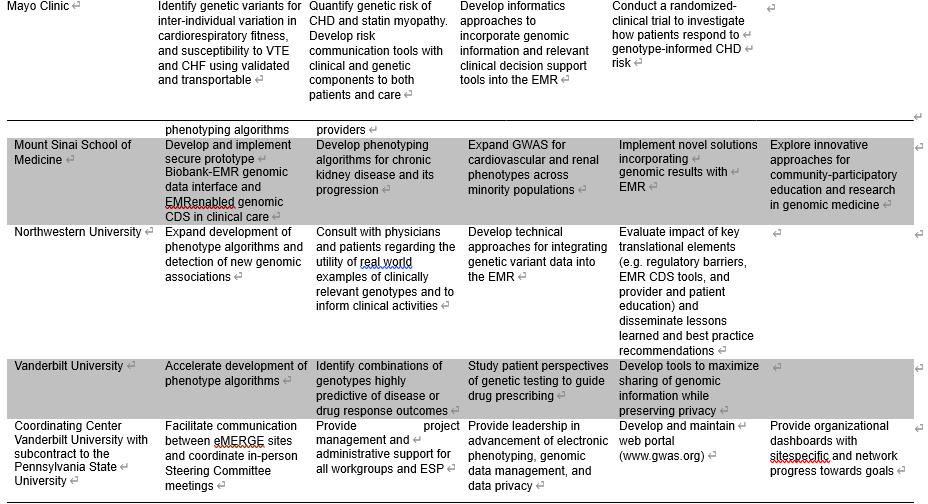
3. Schildcrout, J. S.等人。通过药物遗传学优化药物结果:一个先发制人的基因分型的例子。Clin Pharmacol & Therap 2012; 92:235-242。

## 协调中心

eMERGE网络协调中心(CC)位于范德比尔特大学医学中心，是eMERGE的中心通信和管理中心。协调中心协调所有亲自指导委员会会议，并向所有六个网络工作组和外部科学小组提供项目管理和行政支持。CC已经开发并维护了主要的网络门户(www.gwas.org)，并提供了组织指示板，详细描述了特定于站点和针对目标的网络进度。该中心还在三个关键领域发挥领导作用:电子表型的发展、基因组数据管理和数据隐私。表型工作包括开发基于web的电子表型库(http://PheKB.org)和开发并实现结合EMR和遗传数据的高通量表型-基因型关联发现方法(PheWAS)。GWAS数据来自超过18000个eMERGE-I受试者，另外24000个eMERGE-II受试者。已开发了一种便于使用的软件来解除身份和保护隐私，所有网站均可使用。

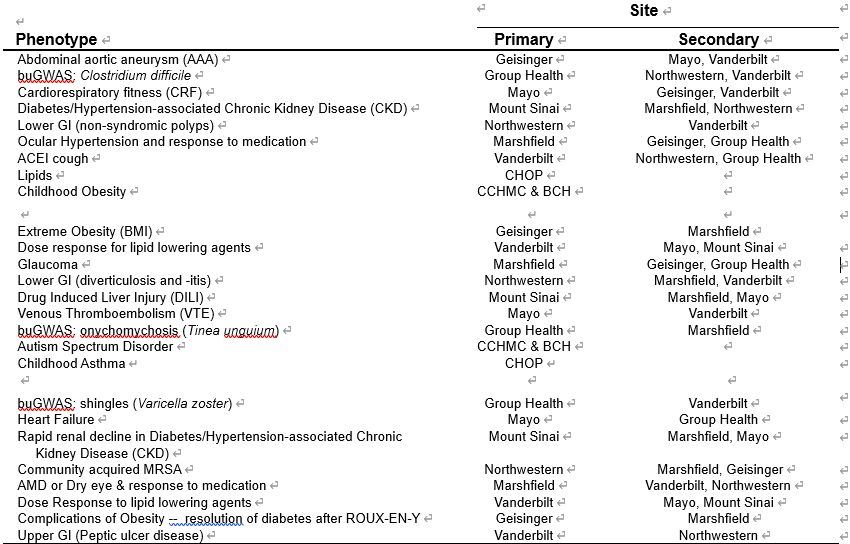
表S1。每个场址紧急情况二的活动摘要





腹主动脉瘤;多动症，注意缺陷多动障碍;CDS，临床决策支持;冠心病;慢性心力衰竭;电子病历;专门委员会，外部科学小组;GERD，胃食管反流病;GWAS，全基因组关联研究;PheWAS，全酚关联研究;静脉血栓栓塞,静脉tromboembolism。有关生物反应器和EMR的详细信息，请参阅主要部分的表1。

表S2。出现II期表型



eMERGE网络的表现型工作组的目标是跨eMERGE的表现型算法的创建、验证和执行。在每个站点上采用和执行算法来识别每个感兴趣的表现型的病例和控制。这些表现型算法通常由账单代码、实验室和测试结果、药物暴露和自然语言处理的组合组成。这些算法通常需要6-12个月的时间，通过一个设计和迭代验证的过程来开发。跨多个站点的验证是开发可移植算法的关键，该算法可用于所有新兴站点和网络外部。