



Inova Genomes Publication List

Identification of copy number variants in whole-genome data using Reference Coverage Profiles

Glusman G, Severson A, Dhankani V, Robinson M, Farrah T, Mauldin DE, Stittrich AB, Ament SA, Roach JC, Brunkow ME, Bodian DL, Vockley JG, Shmulevich I, Niederhuber JE, Hood L.

Front Genet. Feb 17; 6:45, 2015

The identification of DNA copy numbers from short-read sequencing data remains a challenge for both technical and algorithmic reasons. The raw data for these analyses are measured in tens to hundreds of gigabytes per genome; transmitting, storing, and analyzing such large files is cumbersome, particularly for methods that analyze several samples simultaneously. We developed a very efficient representation of depth of coverage (150-1000x compression) that enables such analyses. Current methods for analyzing variants in whole-genome sequencing (WGS) data frequently miss copy number variants (CNVs), particularly hemizygous deletions in the 1-100 kb range. To fill this gap, we developed a method to identify CNVs in individual genomes, based on comparison to joint profiles pre-computed from a large set of genomes. We analyzed depth of coverage in over 6000 high quality (>40x) genomes. The depth of cover-

age has strong sequence-specific fluctuations only partially explained by global parameters like %GC. To account for these fluctuations, we constructed multi-genome profiles representing the observed or inferred diploid depth of coverage at each position along the genome. These Reference Coverage Profiles (RCPs) take into account the diverse technologies and pipeline versions used. Normalization of the scaled coverage to the RCP followed by hidden Markov model (HMM) segmentation enables efficient detection of CNVs and large deletions in individual genomes. Use of pre-computed multi-genome coverage profiles improves our ability to analyze each individual genome. We make available RCPs and tools for performing these analyses on personal genomes. We expect the increased sensitivity and specificity for individual genome analysis to be critical for achieving clinical-grade genome interpretation.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/25741365>

Expanding the Phenotypic Spectrum in EP300-Related Rubinstein-Taybi Syndrome

Solomon BD, Bodian DL, Khromykh A, Gomez Mora G, Lanpher BL, Iyer RK, Baveja R, Vockley JG, Niederhuber JE.

Am J Med Genet A. May;167(5):1111-6, 2015

Rubinstein-Taybi syndrome (RSTS) can be caused by heterozygous mutations or deletions involving CREBBP or, less commonly, EP300. To date, only 15 patients with EP300 mutations have been clinically described. Frequently reported manifestations in these patients include characteristic facial and limb features, varying degrees of neurocognitive dysfunction, and maternal preeclampsia. Other congenital anomalies are less frequently reported. We describe a child found to have a de novo EP300 mutation (c.4933C>T, predicted to result in p.Arg1645X) through research-based whole-genome sequencing of the family trio. The child's presentation involved dysmorphic features as well as unilateral renal agenesis, a myelomeningocele, and minor genitouri-

nary anomalies. The involvement of congenital anomalies in all 16 clinically described patients with EP300 mutations (25% of which have been identified by "hypothesis free" methods, including microarray, exome, and whole-genome sequencing) is reviewed. In summary, genitourinary anomalies have been identified in 38%, cardiovascular anomalies in 25%, spinal/vertebral anomalies in 19%, other skeletal anomalies in 19%, brain anomalies in 13%, and renal anomalies in 6%. Our patient expands the phenotypic spectrum in EP300-related RSTS; this case demonstrates the evolving practice of clinical genomics related to increasing availability of genomic sequencing methods.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/?term=25712426>

Diagnosis of an imprinted-gene syndrome by a novel bioinformatics analysis of whole-genome sequences from a family trio

Bodian DL, Solomon, BD, Khromykh A, Thach DC, Iyer RK, Link K, Banker RL, Baveja R, Vockley JG, Niederhuber JE.

Mol Genet Genomic Med. Nov;2(6):30-8, 2014

Whole-genome sequencing and whole-exome sequencing are becoming more widely applied in clinical medicine to help diagnose rare genetic diseases. Identification of the underlying causative mutations by genome-wide sequencing is greatly facilitated by concurrent analysis of multiple family members, most often the mother-father-proband trio, using bioinformatics pipelines that filter genetic variants by mode of inheritance. However, current pipelines are limited to Mendelian inheritance patterns and do not specifically address disorders caused by mutations in imprinted genes, such as forms of Angelman syndrome and Beckwith-Wiedemann syndrome. Using publicly available tools, we implemented a genetic inheritance search mode to identify imprinted-

gene mutations. Application of this search mode to whole-genome sequences from a family trio led to a diagnosis for a proband for whom extensive clinical testing and Mendelian inheritance-based sequence analysis were nondiagnostic. The condition in this patient, IMAGE syndrome, is likely caused by the heterozygous mutation c.832A>G (p.Lys278Glu) in the imprinted gene CDKN1C. The genotypes and disease status of six members of the family are consistent with maternal expression of the gene, and allele-biased expression was confirmed by RNA-Seq for the heterozygotes. This analysis demonstrates that an imprinted-gene search mode is a valuable addition to genome sequence analysis pipelines for identifying disease-causative variants.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/?term=25614875>

Mutations in NOTCH1 cause Adams-Oliver syndrome

Stittrich AB, Lehman A, Bodian DL, Ashworth J, Zong Z, Li H, Lam P, Khromykh A, Iyer RK, Vockley JG, Baveja R, Silva ES, Dixon J, Leon EL, Solomon BD, Glusman G, Niederhuber JE, Roach JC, Patel MS

Am J Hum Genet. 2014;95(3):275-84

Rubinstein-Taybi syndrome (RSTS) can be caused by heterozygous mutations or deletions involving CREBBP or, less commonly, EP300. To date, only 15 patients with EP300 mutations have been clinically described. Frequently reported manifestations in these patients include characteristic facial and limb features, varying degrees of neurocognitive dysfunction, and maternal preeclampsia. Other congenital anomalies are less frequently reported. We describe a child found to have a de novo EP300 mutation (c.4933C>T, predicted to result in p.Arg1645X) through research-based whole-genome sequencing of the family trio. The child's presentation involved dysmorphic features as well as unilateral renal agenesis, a myelomeningocele, and minor genitouri-

nary anomalies. The involvement of congenital anomalies in all 16 clinically described patients with EP300 mutations (25% of which have been identified by "hypothesis free" methods, including microarray, exome, and whole-genome sequencing) is reviewed. In summary, genitourinary anomalies have been identified in 38%, cardiovascular anomalies in 25%, spinal/vertebral anomalies in 19%, other skeletal anomalies in 19%, brain anomalies in 13%, and renal anomalies in 6%. Our patient expands the phenotypic spectrum in EP300-related RSTS; this case demonstrates the evolving practice of clinical genomics related to increasing availability of genomic sequencing methods.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/?term=25132448>

Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing

Bodian DL, McCutcheon JN, Kothiyal P, Huddleston KC, Iyer RK, Vockley JG, Niederhuber JE.

PLoS One. 2014 Apr 11;9(4):e94554. doi: 10.1371/journal.pone.0094554. eCollection 2014

Technological advances coupled with decreasing costs are bringing whole genome and whole exome sequencing closer to routine clinical use. One of the hurdles to clinical implementation is the high number of variants of unknown significance. For cancer-susceptibility genes, the difficulty in interpreting the clinical relevance of the genomic variants is compounded by the fact that most of what is known about these variants comes from the study of highly selected populations, such as cancer patients or individuals with a family history of cancer. The genetic variation in known cancer-susceptibility genes in the general population has not been well characterized to date. To address this gap, we profiled the nonsynonymous genomic variation in 158 genes causally implicated in carcinogenesis using high-quality whole genome sequences from an ancestrally

diverse cohort of 681 healthy individuals. We found that all individuals carry multiple variants that may impact cancer susceptibility, with an average of 68 variants per individual. Of the 2,688 allelic variants identified within the cohort, most are very rare, with 75% found in only 1 or 2 individuals in our population. Allele frequencies vary between ancestral groups, and there are 21 variants for which the minor allele in one population is the major allele in another. Detailed analysis of a selected subset of 5 clinically important cancer genes, BRCA1, BRCA2, KRAS, TP53, and PTEN, highlights differences between germline variants and reported somatic mutations. The dataset can serve as a resource of genetic variation in cancer-susceptibility genes in 6 ancestry groups, an important foundation for the interpretation of cancer risk from personal genome sequences.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/?term=24728327>

Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates

Bodian DL, Klein E, Iyer RK, Wong WS, Kothiyal P, Stauffer D, Huddleston KC, Gaither AD, Remsburg I, Khromykh A, Baker RL, Maxwell GL, Vockley JG, Niederhuber JE, Solomon BD1

Genet Med. 2015 Sep 3. doi: 10.1038/gim.2015.111.

PURPOSE:

To assess the potential of whole-genome sequencing (WGS) to replicate and augment results from conventional blood-based newborn screening (NBS).

METHODS:

Research-generated WGS data from an ancestrally diverse cohort of 1,696 infants and both parents of each infant were analyzed for variants in 163 genes involved in disorders included or under discussion for inclusion in US NBS programs. WGS results were compared with results from state NBS and related follow-up testing.

RESULTS:

NBS genes are generally well covered by WGS. There is a median of one (range: 0-6) database-annotated pathogenic variant in the

NBS genes per infant. Results of WGS and NBS in detecting 28 state-screened disorders and four hemoglobin traits were concordant for 88.6% of true positives (n = 35) and 98.9% of true negatives (n = 45,757). Of the five infants affected with a state-screened disorder, WGS identified two whereas NBS detected four. WGS yielded fewer false positives than NBS (0.037 vs. 0.17%) but more results of uncertain significance (0.90 vs. 0.013%).

CONCLUSION:

WGS may help rule in and rule out NBS disorders, pinpoint molecular diagnoses, and detect conditions not amenable to current NBS assays.

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/26334177>

Diagnosis of D-Bifunctional Protein Deficiency through Whole-Genome Sequencing: Implications for Cost-Effective Care

Khromykh A, Solomon BD, Bodian DL, Leon EL, Iyer RK, Baker RL, Ascher DP, Baveja R, Vockley JG, Niederhuber JE

Mol Syndromol 2015;6:141-146

D-Bifunctional protein deficiency, caused by recessive mutations in HSD17B4, is a severe disorder of peroxisomal fatty acid oxidation. Nonspecific clinical features may contribute to diagnostic challenges. We describe a newborn female with infantile-onset seizures and nonspecific mild dysmorphisms who underwent extensive genetic workup that resulted in the detection of a novel homozygous mutation (c.302+1_4delGTGA) in the HSD17B4 gene, consistent with a diagnosis of D-bifunctional protein deficiency. By comparing the standard clinical workup to diagnostic analysis per-

formed through research-based whole-genome sequencing (WGS), which independently identified the causative mutation, we demonstrated the ability of genomic sequencing to serve as a timely and cost-effective diagnostic tool for the molecular diagnosis of apparent and occult newborn diseases. As genomic sequencing becomes more available and affordable, we anticipate that WGS and related omics technologies will eventually replace the traditional tiered approach to newborn diagnostic workup.