

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 718 343**

51 Int. Cl.:

C12Q 1/6883 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.11.2013 PCT/US2013/069015**

87 Fecha y número de publicación internacional: **15.05.2014 WO14074755**

96 Fecha de presentación y número de la solicitud europea: **07.11.2013 E 13792839 (6)**

97 Fecha y número de publicación de la concesión europea: **09.01.2019 EP 2917369**

54 Título: **Evaluación del riesgo de discapacidad lectora y lingüística**

30 Prioridad:

07.11.2012 US 201261723774 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

01.07.2019

73 Titular/es:

**YALE UNIVERSITY (100.0%)
Two Whitney Avenue
New Haven, CT 06511, US**

72 Inventor/es:

GRUEN, JEFFREY, R.

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 718 343 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Evaluación del riesgo de discapacidad lectora y lingüística

Antecedentes

5 Las dificultades del aprendizaje (DA) específicas son trastornos caracterizados por una dificultad inesperada con un modo de aprendizaje específico, a pesar de la presencia de un coeficiente intelectual (CI) adecuado y oportunidades educativas adecuadas. Las DA pueden afectar las competencias en las áreas de la lectura, la escritura, el habla y las matemáticas, entre otras, pero las más comunes se relacionan con el lenguaje. El Instituto Nacional de Salud y Desarrollo Infantil (NICHD) estima que entre el 15 y el 20 % de los estadounidenses tiene una DA asociada al lenguaje, y que es la discapacidad lectora (DL) la que afecta a la mayoría (1). La DL, también conocida como dislexia, es una dificultad específica con el procesamiento del lenguaje escrito (2). Otra DA, el trastorno del lenguaje (TL), se caracteriza por una dificultad para procesar y expresar el lenguaje hablado (3). Estas DA son con frecuencia concomitantes; los niños con TL tienen un mayor riesgo de desarrollar DL (3). Debido a que las competencias lectoras y lingüísticas son fundamentales para lograr el éxito académico, las personas afectadas corren el riesgo de sufrir efectos psicológicos adversos, así como también de tener posibilidades educativas y laborales limitadas (2).
 10 Además, la prevalencia de estas DA hace que el costo de las medidas de apoyo sea una carga para el sistema educativo (4). La intervención es más eficaz cuanto antes se administre (2), lo que hace que la detección temprana de las personas de alto riesgo sea una posibilidad deseable.

El documento WO 2006/032021 describe un haplotipo de siete marcadores (AGCTAGA), que abarca los exones de 5 a 8 de DCDC2, que se correlaciona con la dislexia.

20 HAROLD *et al.* ("Evidencia adicional de que el gen KIAA0319 confiere susceptibilidad a la dislexia del desarrollo", MOLECULAR PSYCHIATRY, 2006, vol. 11, n.º 12, páginas 1085-1091), SCERRI *et al.* ("DCDC2, KIAA0319 y CMIP están asociados a rasgos relacionados con la lectura", BIOLOGICAL PSYCHIATRY, vol. 70, n.º 3, 2011, páginas 237-245), COPE *et al.* ("Las variantes en el locus DYX2 están asociadas a la activación cerebral alterada en regiones cerebrales relacionadas con la lectura en sujetos con discapacidad lectora", NEUROIMAGE, 2012, vol. 63, n.º 1, páginas 148-156) y LIND *et al.* ("Dislexia y DCDC2: la variación normal en la lectura y la ortografía está asociada a polimorfismos de DCDC2 en una muestra de población australiana", EUROPEAN JOURNAL OF HUMAN GENETICS, 2010, vol. 18, n.º 6, páginas 668-673) describen el DCDC2 como un gen marcador asociado a las dificultades del aprendizaje y describen varios polimorfismos que están relacionados con diferentes discapacidades lectoras y cognitivas.

30 Compendio

La presente divulgación proporciona métodos, matrices y micromatrices tal como se definen en las reivindicaciones adjuntas. Tal como se describe en el presente documento, dos haplotipos, ambos en el mismo bloque de haplotipos de seis marcadores en el gen de riesgo de discapacidad lectora (DL), DCDC2, están asociados, respectivamente, a la discapacidad lectora y al trastorno del lenguaje (TL). Cada uno de los haplotipos presenta un desequilibrio de enlace muy fuerte con un alelo de BV677278 (también conocido como READ1), que es una repetición corta en tándem STR de compuestos polimórficos asociada a la discapacidad lectora y capaz de modular la expresión del promotor de DCDC2. Se ha demostrado que BV677278 (READ1) se une específicamente a ETV6, un potente regulador transcripcional y protooncogén, *in vivo*. BV677278 se une a la proteína nuclear expresada en el cerebro con una especificidad muy alta y es capaz de modular la expresión del gen reportero del promotor de DCDC2 de una manera específica con respecto a los alelos. Los patrones de activación en las áreas del cerebro relacionadas con la lectura, medidos por imágenes de resonancia magnética funcional, están influenciados por los alelos de BV677278. El trabajo que se describe en el presente muestra que BV677278 está asociado a la lectura y el lenguaje, y que al menos dos alelos de BV677278 tienen un efecto perjudicial sobre la lectura y el lenguaje. El alelo 5 es importante para la dislexia (DL) y el alelo 6 es importante para el trastorno del lenguaje. BV677278 ha sufrido un cambio de nombre y en el presente también se denomina READ1, que significa "elemento regulador asociado a la dislexia 1".
 45 Los dos términos se usan indistintamente.

Como también se describe en el presente documento, los haplotipos o alelos de riesgo de DCDC2 interactúan con un haplotipo de riesgo de KIAA0319 de una manera sinérgica. La sinergia entre los alelos de BV677278 (READ1) o los haplotipos de riesgo de DCDC2 y los haplotipos de KIAA0319 en la disminución del rendimiento en la eliminación de fonemas (muy importante para la lectura), la ortografía, el coeficiente intelectual total y el coeficiente intelectual de rendimiento no se ha descrito anteriormente. Juntos, el efecto es de 3 a 8 veces mayor que si está presente la versión perjudicial de un alelo o un haplotipo de riesgo.

En particular, en el presente se describe un bloque de haplotipos de seis marcadores dentro de DCDC2, de los cuales dos haplotipos (CGCGAG y GACGAG) están asociados a un rendimiento muy bajo en una tarea de eliminación de fonemas y una medición compuesta del lenguaje, respectivamente. Los dos haplotipos muestran una fuerte asociación con sus respectivos fenotipos: CGCGAG con DL y GACGAG con TL. Los portadores del haplotipo CGCGAG, en promedio, mostraron un rendimiento significativamente inferior en ocho mediciones relacionadas con

la lectura, en comparación con los no portadores, y los portadores del haplotipo GACGAG mostraron un rendimiento medio significativamente inferior en la medición compuesta del lenguaje WOLD/WR.

En el presente documento se describen además los métodos para identificar o ayudar a identificar a un individuo que corre el riesgo de desarrollar al menos una (o más) dificultad(es) del aprendizaje (DA), para los cuales se analiza una muestra obtenida del individuo a fin de determinar la presencia de al menos un haplotipo (o más) en el locus DYX2 (cromosoma 6p; 6p22) que está asociado a la susceptibilidad de desarrollar al menos una (o más) DA en seres humanos. La presencia de al menos un haplotipo que esté asociado a la susceptibilidad de desarrollar al menos una (o más) DA en seres humanos en el locus DYX2 indica que el individuo corre el riesgo de desarrollar al menos una (o más) DA. En algunos aspectos de la divulgación, dicho al menos una DA es la discapacidad lectora (DL) o el trastorno del lenguaje (TL). Dicho al menos un haplotipo puede ubicarse en el gen DCDC2, dentro del locus DYX2, o en el gen KIAA0319, dentro del locus DYX2. De manera alternativa, se analiza una muestra para determinar la presencia de un haplotipo ubicado en el gen DCDC2, dentro del locus DYX2, y la presencia de un haplotipo ubicado en el gen KIAA0319, dentro del locus DYX2. Dicho al menos un haplotipo puede comprender (a) CGCGAG, GACGAG, o ambos en un gen DCDC2 dentro del locus DYX2; (b) uno o más polimorfismos de un solo nucleótido (SNP) asociados a una variante de KIAA0319, como rs4504469, rs2038137 o rs2143340 o cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340; o (c) cualquier combinación de los haplotipos en (a) y (b). En el presente documento también se describen métodos para tratar a un individuo con una DA sospechada o confirmada. El tratamiento puede incluir, por ejemplo, inhibir a ETV6 en el individuo y/o proporcionar servicios diseñados para abordar o remediar ciertos aspectos asociados con una DA, como la DL o el TL, tal como proporcionar intervención, incluidos servicios y materiales como, entre otros, utilizar técnicas especiales de enseñanza; realizar modificaciones en el aula, como proporcionar tiempo adicional para realizar las tareas y pruebas grabadas para permitir que el individuo escuche, en lugar de leer, las pruebas; utilizar libros grabados; utilizar programas de procesamiento de texto con características de verificación ortográfica; ayudar al individuo a aprender a través de experiencias multisensoriales; enseñar herramientas de afrontamiento; y proporcionar servicios para fortalecer la capacidad del individuo para reconocer y pronunciar palabras.

En el presente documento se describe un método para determinar si una muestra obtenida de un individuo comprende ácido nucleico que comprende un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos, que comprende analizar una muestra que comprende ácido nucleico del individuo para determinar la presencia en el locus DYX2 de al menos uno de los siguientes marcadores: (a) CGCGAG en un gen DCDC2; (b) GACGAG en un gen DCDC2; (c) tanto CGCGAG como GACGAG en un gen DCDC2; (d) rs4504469 en un gen KIAA0319; (e) rs2038137 en un gen KIAA0319; (f) rs2143340 en un gen KIAA0319; (g) cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319; y (h) cualquier combinación de CGCGAG en un gen DCDC2; GACGAG en un gen DCDC2; tanto CGCGAG como GACGAG en un gen DCDC2; rs4504469 en un gen KIAA0319; rs2038137 en un gen KIAA0319; rs2143340 en un gen KIAA0319; y cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319, en donde si la muestra comprende al menos uno de (a) - (h), la muestra comprende un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje en seres humanos. En un aspecto, la muestra se analiza para determinar la presencia de al menos un marcador de (a), (b) y (c), y al menos un marcador de (d), (e), (f) y (g). En varios aspectos de la divulgación, la muestra se analiza para determinar la presencia de dos o más marcadores de (a), (b) y (c), y dos o más marcadores de (d), (e), (f) y (g); solo marcadores de (a), (b) y (c); solo marcadores de (d), (e), (f) y (g); o cualquier combinación de los marcadores de (a) - (h). La DA puede ser DL o TL, o tanto DL como TL. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos. En un aspecto, el método comprende además analizar la muestra para determinar la presencia del alelo 5 del gen DCDC2, como se presenta en este documento (por ejemplo, la SEQ ID NO: 35); el alelo 6 del gen DCDC2, como se presenta en este documento (por ejemplo, SEQ ID NO: 36), o ambos.

Otro aspecto de la divulgación es un método para analizar una muestra en busca de un marcador de un haplotipo asociado a la susceptibilidad de desarrollar una dificultad de aprendizaje (DA) en seres humanos, que comprende: (a) obtener una muestra que comprenda ácido nucleico de un individuo; y (b) determinar si la muestra comprende al menos uno de los siguientes: (i) CGCGAG en un gen DCDC2 en el locus DYX2; (ii) GACGAG en un gen DCDC2 en el locus DYX2; (iii) tanto CGCGAG como GACGAG en un gen DCDC2 en el locus DYX2; (iv) rs4504469 en un gen KIAA0319 en el locus DYX2; (v) rs2038137 en un gen KIAA0319 en el locus DYX2; (vi) rs2143340 en un gen KIAA0319 en el locus DYX2; (vii) cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319 en el locus DYX2; y (viii) cualquier combinación de CGCGAG en un gen DCDC2; GACGAG en un gen DCDC2 en el locus DYX2; tanto CGCGAG como GACGAG en un gen DCDC2 en el locus DYX2; rs4504469 en un gen KIAA0319 en el locus DYX2; rs2038137 en un gen KIAA0319 en el locus DYX2; rs2143340 en un gen KIAA0319 en el locus DYX2; y cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319 en el locus DYX2, en donde si la muestra comprende al menos un marcador de (i) - (viii), la muestra comprende un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje en seres humanos.

En un aspecto, la muestra se analiza para detectar al menos un marcador de (i), (ii) y (iii) y al menos un marcador de (iv), (v), (vi) y (vii). La DA es discapacidad lectora (DL) o trastorno del lenguaje (TL), o tanto DL como TL. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos.

Un aspecto adicional de la divulgación es un método para determinar si una muestra obtenida de un individuo comprende ácido nucleico que comprende un alelo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos, que comprende: analizar una muestra que comprende ácidos nucleicos del individuo para determinar la presencia del alelo 5 del gen DCDC2 en el locus DYX2 (por ejemplo, SEQ ID NO: 35), alelo 6 del gen DCDC2 en el locus DYX2 (por ejemplo, SEQ ID NO: 36), o tanto el alelo 5 del gen DCDC2 como el alelo 6 del gen DCDC2, en donde si la muestra comprende al menos uno del alelo 5 y el alelo 6, la muestra comprende un alelo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje. La DA es discapacidad lectora (DL) o trastorno del lenguaje (TL), o tanto DL como TL. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos.

Un aspecto adicional de la divulgación es un método para determinar si una muestra obtenida de un individuo comprende al menos un marcador asociado a discapacidad lectora (DL) y trastorno del lenguaje (TL) concomitantes en seres humanos, que comprende: (a) obtener una muestra que contenga ácido nucleico del individuo y (b) analizar la muestra para determinar la presencia de (i) al menos un marcador en el gen DCDC2 en el locus DYX2 para un haplotipo asociado a la susceptibilidad de desarrollar DA en seres humanos y (ii) al menos un marcador en el gen KIAA0319 en el locus DXY2 para un haplotipo asociado con la susceptibilidad de desarrollar TL en seres humanos, en donde si la muestra comprende un marcador de (i) y un marcador de (ii), la muestra comprende al menos un marcador asociado a DL y TL concomitantes. El marcador de (b) (i) y el marcador de (b) (ii) pueden ser el mismo marcador o dos marcadores diferentes. En un aspecto, dicho al menos un marcador de (b)(i) es CGCGAG o GACGAG, y dicho al menos un marcador de (b)(ii) es rs4504469; rs2038137; o rs2143340. En cualquiera de estos aspectos de la divulgación, el método puede comprender además analizar la muestra para determinar la presencia del alelo 5 del gen DCDC2 en el locus DYX2, el alelo 6 del gen DCDC2 del locus DYX2, o tanto el alelo 5 del gen DCDC2 en el locus DYX2 como el alelo 6 del gen DCDC2 en el locus DYX2. En aún otros aspectos de la divulgación, dicho al menos un marcador de (b)(i) y dicho al menos un marcador de (b)(ii) se seleccionan de: rs12636438; rs1679255; rs9521789; rs1983931; rs9814232; rs7995158; rs6573225; rs4082518; rs442555; rs259521; rs16889556; rs1047782; rs1530680; rs12667130; rs6965855; rs985080; rs4726782; rs1718101; rs10487689; rs1918296; rs737533; rs4504469; rs2038137; rs2143340; rs9295626; rs7763790 rs6935076; rs2817201; rs10456309; rs4576240; rs17307478; rs9356939; rs7763790; rs6456621; rs6456624; rs6935076; rs2038137; rs3756821; rs1883593; rs3212236; rs6456621; rs12193738; rs2817198; rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; y rs807694. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un marcador asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos.

Otro aspecto de la divulgación es un método para determinar si una muestra obtenida de un individuo comprende un marcador asociado a trastorno del lenguaje (TL) en seres humanos, lo que comprende analizar una muestra obtenida del individuo para determinar la presencia de al menos uno de los siguientes marcadores: GACGAG en un gen DCDC2 en el locus DYX2; rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; rs12193738; rs2817198; rs10456309; rs985080; rs1554690; rs2533096; rs6951437; rs344470; rs344468; rs807694; rs482700; rs7695228; rs1940309; rs505277; rs476739; rs867036; rs867035; rs2071674; rs7694946; rs4823324; y un marcador para al menos uno de los siguientes genes: NEK2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; y PGD. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un marcador asociado a la susceptibilidad de desarrollar TL en seres humanos.

Un aspecto adicional de la divulgación es un método para determinar si una muestra obtenida de un individuo comprende un marcador asociado a la discapacidad lectora (DL) en seres humanos, lo que comprende analizar una muestra obtenida del individuo para determinar la presencia de al menos uno de los siguientes marcadores:

CGCGAG en un gen DCDC2 en el locus DYX2; rs180950; rs2590673; rs892100; rs1792745; rs12546767; rs12634033; rs892270; rs10887149; rs10041417; rs6792971; rs4725745; rs12444778; rs1444186; rs2294691; rs10456309; rs1562422; y un marcador para al menos uno de los siguientes genes: MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622; OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de un marcador asociado a la susceptibilidad de desarrollar DL en seres humanos.

Otro aspecto de la divulgación es un método para determinar si los ácidos nucleicos (ADN, ARN) en un individuo comprenden marcadores de haplotipos que interactúen de una manera sinérgica y produzcan una dificultad del aprendizaje (DA) en seres humanos, que comprende: (a) obtener una muestra que comprenda ácidos nucleicos de un individuo y (b) analizar la muestra para determinar la presencia de al menos un haplotipo de riesgo de DCDC2 o un alelo de riesgo de DCDC2 y al menos un haplotipo de riesgo de KIAA0319, en donde dicho al menos un haplotipo de riesgo de DCDC2 es CGCGAG o GACGAG, dicho al menos un alelo de riesgo de DCDC2 es el alelo 5 del gen DCDC2 en el locus DYX2 (SEQ ID NO: 35) o el alelo 6 del gen DCDC2 en el locus DYX2 (SEQ ID NO: 36), y dicho al menos un haplotipo de riesgo de KIAA0319 es una variante de haplotipo de KIAA0319 que comprende un SNP que es rs4504469; rs2038137; o rs2143340, y en donde si la muestra comprende al menos un haplotipo de riesgo de DCDC2 o al menos un alelo de riesgo de DCDC2 y al menos un haplotipo de riesgo de KIAA0319, los ácidos nucleicos comprenden marcadores de haplotipos que interactúan de forma sinérgica y producen una DA en seres humanos. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar ADN genómico y determinar la presencia o ausencia de marcadores que interactúen de una manera sinérgica y produzcan DA en seres humanos.

Un aspecto adicional de la divulgación es un método para identificar o ayudar a identificar a un individuo que corre el riesgo de desarrollar al menos una dificultad del aprendizaje (DA), que comprende analizar una muestra obtenida del individuo para determinar la presencia en el locus DYX2 de al menos un haplotipo asociado a la susceptibilidad de desarrollar una DA en seres humanos, en donde la presencia en el locus DYX2 de al menos un haplotipo asociado a la susceptibilidad de desarrollar una DA en seres humanos indica que el individuo corre el riesgo de desarrollar una DA. Al menos una DA es una discapacidad lectora o trastorno del lenguaje. Dicho al menos un haplotipo está localizado en el gen DCDC2 dentro del locus DYX2 o en el gen KIAA0319 dentro del locus DYX2. Dicho al menos un haplotipo puede comprender (a) CGCGAG o GACGAG en un gen DCDC2 dentro del locus DYX2; o (b) rs4504469, rs2038137, rs2143340, o cualquier combinación de estos en un gen KIAA0319 dentro del locus DYX2; o (c) cualquier combinación de los haplotipos en (a) y (b). Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un haplotipo de interés; los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar el ADN genómico y determinar la presencia o ausencia en el locus DYX2 de al menos un haplotipo que esté asociado a la susceptibilidad de desarrollar una DA en seres humanos. La presencia en el locus DYX2 de al menos un haplotipo que esté asociado a la susceptibilidad de desarrollar una DA en seres humanos indica que el individuo corre el riesgo de desarrollar una DA.

También se describe en el presente documento un método para determinar si una muestra obtenida de un individuo comprende un marcador de susceptibilidad de desarrollar una dificultad del aprendizaje (DA) que es discapacidad lectora (DL) o trastorno del lenguaje (TL), que comprende: obtener una muestra que comprenda ácidos nucleicos del individuo y determinar si la muestra comprende al menos un marcador seleccionado del grupo que consiste en: rs12636438; rs1679255; rs9521789; rs1983931; rs9814232; rs7995158; rs6573225; rs4082518; rs442555; rs259521; rs482700; rs7695228; rs1940309; rs505277; rs476739; rs867036; rs867035; rs2071674; rs7694946; rs4823324; rs180950; rs2590673; rs892100; rs1792745; rs12546767; rs12634033; rs892270; rs10887149; rs10041417; rs6792971; rs12636438; rs1679255; rs9521789; rs476739; rs505277; rs482700; rs7695228; rs867036; rs867035; rs1940309; rs16889556; rs1047782; rs1530680; rs12667130; rs6965855; rs985080; rs4726782; rs1718101; rs10487689; rs1918296; rs737533; rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; rs12193738; rs2817198; rs10456309; rs985080; rs1554690; rs2533096; rs6951437; rs344470; rs344468; rs4725745; rs12444778; rs1444186; rs2294691; rs10456309; rs1562422; rs807694; rs3756814; rs3777663; rs9295626; rs7763790; rs6935076; rs9348646; rs2328791; rs2328791; rs2817201, rs9295626; rs4576240; rs17307478; rs9356939; rs7763790; rs6456621; rs6456624, rs6935076, rs2038137, rs3756821, rs1883593, rs3212236; rs3777663, rs3756814, rs6931809, rs6916186, rs6933328, rs17491647; rs2328791; rs33914824a; rs807694a; rs707864a; rs10456301a; rs16889066a; rs9379651a; rs2817201; rs9295626; rs10456309; rs4576240; rs17307478; rs9356939; rs7763790; rs6456621; rs3756821; rs1883593; rs3212236; rs2294691; rs3777663; rs3756814; rs6931809; rs6916186; rs6933328; rs17491647; rs9348646; rs1562422 y un marcador para cada uno de los siguientes genes: R5H2; OR5H6; RRAGA; OR6B3; UMOD; A26C1A; FAM29A; CHRNA1; IFIT5; LOC643905; K2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; PGD; MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622;

OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF. La DA es discapacidad lectora (DL) o trastorno del lenguaje (TL), o tanto DL como TL. Se puede analizar cualquier muestra que contenga ácido nucleico (por ejemplo, ADN genómico; ARN) que pueda analizarse en busca de un marcador de susceptibilidad de desarrollar una dificultad del aprendizaje (DA) que sea discapacidad lectora (DL) o trastorno del lenguaje (TL); los métodos para analizar ácidos nucleicos son bien conocidos en la técnica y también se describen en el presente documento. Incluyen, entre otros, métodos mediados por hibridación y secuenciación. La muestra puede ser, por ejemplo, sangre, células o tejido. De manera alternativa, se puede secuenciar el ADN genómico y determinar la presencia o ausencia de dicho marcador o marcadores.

En el presente documento se describe además un método para identificar o ayudar a identificar a un individuo que corre el riesgo de desarrollar al menos una dificultad del aprendizaje (DA), que comprende analizar una muestra obtenida del individuo para determinar la presencia en el locus DYX2 de al menos un haplotipo asociado a la susceptibilidad de desarrollar una DA en seres humanos, en donde la presencia en el locus DYX2 de al menos un haplotipo que está asociado a la susceptibilidad de desarrollar una DA en seres humanos indica que el individuo corre el riesgo de desarrollar una DA. Al menos una DA es una discapacidad lectora (DL) o un trastorno del lenguaje (TL). Dicho al menos un haplotipo se encuentra en el gen DCDC2 dentro del locus DYX2 o en el gen KIAA0319 dentro del locus DYX2, y puede comprender (a) CGCGAG, GACGAG, o tanto CGCGAG como GACGAG, en un gen DCDC2 dentro del locus DYX2; o (b) rs4504469, rs2038137, rs2143340, o cualquier combinación de estos en un gen KIAA0319 dentro del locus DYX2; o (c) cualquier combinación de los haplotipos en (a) y (b). En el método, el ensayo comprende un método mediado por hibridación, secuenciación de ácidos nucleicos, o tanto un método mediado por hibridación como una secuenciación de ácidos nucleicos. La muestra puede ser sangre, células o tejido.

Otro aspecto de la divulgación es un método para identificar a un individuo que tiene, o es susceptible de desarrollar, una dificultad del aprendizaje (DA), que comprende obtener una muestra que comprende ácido nucleico de un individuo; determinar si el ácido nucleico en la muestra comprende un haplotipo del gen DCDC2 en el locus DYX2 asociado a la susceptibilidad de desarrollar una discapacidad lectora (DL) y un haplotipo del gen KIAA0319 asociado a la susceptibilidad de desarrollar un trastorno del lenguaje (TL), en donde el haplotipo del gen DCDC2 y el haplotipo del gen KIAA0319 interactúan de forma sinérgica para disminuir el rendimiento en la eliminación de fonemas y provocar una dificultad del aprendizaje (DA) en seres humanos, en donde si la muestra comprende ambos haplotipos, se identifica que el individuo tiene o es susceptible de desarrollar una DA. En este método, la determinación comprende un método mediado por hibridación, una secuenciación de ácidos nucleicos, o tanto un método mediado por hibridación como una secuenciación de ácidos nucleicos. La muestra es sangre, células o tejido.

En algunos aspectos de la divulgación, un método por el cual se identifica a un individuo que tiene o es susceptible de desarrollar una dificultad del aprendizaje (DA) comprende tratar al individuo así identificado. El tratamiento comprende proporcionar intervenciones, incluidos servicios y materiales como, entre otros, utilizar técnicas especiales de enseñanza; realizar modificaciones en el aula, como proporcionar tiempo adicional para realizar las tareas y pruebas grabadas para permitir que el individuo escuche, en lugar de leer, las pruebas; utilizar libros grabados; utilizar programas de procesamiento de texto con características de verificación ortográfica; ayudar al individuo a aprender a través de experiencias multisensoriales; enseñar herramientas de afrontamiento; y proporcionar servicios para fortalecer la capacidad del individuo para reconocer y pronunciar palabras. Véase, por ejemplo, nichd.nih.gov/health/topics/learning/conditioninfo. ¿Cuáles son los tratamientos para las dificultades del aprendizaje?

Otro aspecto de la divulgación es un método para tratar a un individuo por una dificultad del aprendizaje (DA) que comprende inhibir a ETV6 en el individuo. El individuo tiene DL, TL, o ambos.

En algunos aspectos de la divulgación, un método por el cual se identifica a un individuo que corre el riesgo de desarrollar un trastorno del aprendizaje comprende además monitorear al individuo cuyo riesgo de desarrollar una dificultad del aprendizaje se ha identificado para evaluar si se produce el desarrollo de una dificultad del aprendizaje y, si se produce dicha dificultad, tratar al individuo, en donde el tratamiento comprende proporcionar intervenciones, incluidos servicios y materiales, que incluyen, entre otros, utilizar técnicas especiales de enseñanza; realizar modificaciones en el aula, como proporcionar tiempo adicional para realizar las tareas y pruebas grabadas para permitir que el individuo escuche, en lugar de leer, las pruebas; utilizar libros grabados; utilizar programas de procesamiento de texto con características de verificación ortográfica; ayudar al individuo a aprender a través de experiencias multisensoriales; enseñar herramientas de afrontamiento; y proporcionar servicios a fin de fortalecer la capacidad del individuo para reconocer y pronunciar palabras.

Véase, por ejemplo, nichd.nih.gov/health/topics/learning/conditioninfo. ¿Cuáles son los tratamientos para las dificultades del aprendizaje?

En el presente documento también se describen matrices, tales como micromatrices (matrices de ADN o micromatrices). De acuerdo con un aspecto, se describe una matriz (por ejemplo, una micromatriz) para identificar o ayudar a identificar a un individuo que corre el riesgo de desarrollar al menos una dificultad del aprendizaje (DA). La matriz comprende un soporte que tiene una pluralidad de regiones discretas (por ejemplo, manchas), teniendo cada región discreta (habiéndose adherido a ella) uno o más fragmentos de ácido nucleico (por ejemplo, sondas)

manchados, o unidos o depositados de otro modo en ellas. Normalmente, cada región discreta lleva un reactivo, como ácido nucleico (ADN, ARN) que detecta un marcador (por ejemplo, SNP, marcador de haplotipo, alelo, etc.) asociado a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL) en seres humanos. Los fragmentos de ácido nucleico son complementarios a los ácidos nucleicos (por ejemplo, ADN, como ADN genómico, o ARN, como ARNm) que son marcadores para un gen variante, como variantes de DCDC2, KIAA0319, y otros nombrados en el presente, asociados a la susceptibilidad de desarrollar al menos una DA (por ejemplo, como se describe en el presente). Los fragmentos de ácido nucleico en una región discreta particular pueden ser de cualquier longitud y secuencia (por ejemplo, que complementa el ácido nucleico que comprende un marcador) adecuada para la detección de cualquier marcador descrito en el presente documento. Por ejemplo, en algunos aspectos de la divulgación, un fragmento de ácido nucleico (por ejemplo, sonda, sonda SNP) tiene una longitud de entre 10 y 100 nucleótidos. En algunos aspectos de la divulgación, un fragmento de ácido nucleico tiene entre alrededor de 20 y 80, alrededor de 30 y 60, o alrededor de 40 y 50 nucleótidos (nt) de longitud. En aspectos específicos de la divulgación, las sondas tienen 25 nt, 30 nt, 35 nt o 40 nt de longitud. Véase, por ejemplo, LaFramboise, T., Nucl. Acids Res. (2009) 37 (13): 4181-4193. En algunos aspectos de la divulgación, una región discreta particular comprende una pluralidad de fragmentos de ácido nucleico (por ejemplo, sondas, sondas SNP), cada uno de los cuales es capaz de hibridarse con un marcador particular. En algunos aspectos de la divulgación, la pluralidad de fragmentos de ácido nucleico tienen longitudes y secuencias variables (por ejemplo, como se describe en el presente documento). En algunos aspectos de la divulgación, la matriz detecta dos o más marcadores asociados a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en humanos, en donde los dos o más marcadores comprenden uno o más marcadores en un gen DCDC2 y uno o más marcadores en un gen KIAA0319. En algunos aspectos de la divulgación, dicho uno o más marcadores en un gen DCDC2 se seleccionan de CGCGAG, GACGAG, el alelo 5 de READ1 (SEQ ID NO: 35), el alelo 6 de READ1 (SEQ ID NO: 36), o cualquier combinación de dos, tres o cuatro de CGCGAG, GACGAG, el alelo 5 de READ1 (SEQ ID NO: 35) y el alelo 6 de READ1 (SEQ ID NO: 36). En algunos aspectos de la divulgación, el uno o más marcadores en un gen KIAA0319 se seleccionan de rs4504469, rs2038137, rs2143340, o cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340.

En algunos aspectos de la divulgación, la matriz (por ejemplo, la micromatriz) detecta marcadores asociados a la susceptibilidad de desarrollar trastorno del lenguaje (TL) en seres humanos. En algunos aspectos de la divulgación, la matriz comprende regiones discretas (por ejemplo, regiones discretas que comprenden uno o más fragmentos de ácido nucleico) capaces de detectar marcadores en un gen DCDC2, como GACGAG, el alelo 6 de READ1 (SEQ ID NO: 36), rs793845, rs2799373, rs793862, rs793834, rs2792682, rs807704, rs707864, rs807694, o cualquier combinación de estos. En algunos aspectos de la divulgación, la matriz (además) detecta uno o más marcadores en un gen KIAA0319, como rs12193738, rs2817198, rs10456309, o cualquier combinación de estos. En algunos aspectos de la divulgación, la matriz comprende además una o más regiones discretas que comprenden fragmentos de ácido nucleico localizados en el soporte que detecta uno o más marcadores seleccionados de rs985080, rs1554690, rs2533096, rs6951437, rs344470, rs344468, rs482700, rs7695228, rs1940309, rs505277, rs476739, rs867036, rs867035, rs2071674, rs7694946, rs4823324, y marcadores para los siguientes genes: NEK2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; y PGD.

En algunos aspectos de la divulgación, la matriz (por ejemplo, la micromatriz) detecta marcadores asociados a la susceptibilidad de desarrollar una discapacidad lectora (DL) en seres humanos. En algunos aspectos de la divulgación, la matriz comprende regiones discretas (por ejemplo, regiones discretas que comprenden uno o más fragmentos de ácido nucleico) capaces de detectar marcadores en un gen DCDC2, CGCGAG y el alelo 5 de READ1 (SEQ ID NO: 35), o ambos. En algunos aspectos de la divulgación, la matriz (además) detecta uno o más marcadores en un gen KIAA0319, como rs10456309. En algunos aspectos de la divulgación, la matriz comprende además uno o más regiones discretas que comprenden fragmentos de ácido nucleico manchados sobre el soporte que detectan uno o más marcadores seleccionados de rs180950, rs2590673, rs892100, rs1792745, rs12546767, rs12634033, rs892270, rs10887149, rs10041417, rs6792971, rs4725745, rs12444778, rs1444186, rs2294691, rs10456309, rs1562422 y marcadores para los siguientes genes: MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622; OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF.

En algunos aspectos de la divulgación, se describe una matriz (por ejemplo, micromatriz) que detecta uno o más marcadores asociados con la susceptibilidad de desarrollar un DA en seres humanos, en donde dicho uno o más marcadores se seleccionan de rs12636438; rs1679255; rs9521789; rs1983931; rs9814232; rs7995158; rs6573225; rs4082518; rs442555; rs259521; rs482700; rs7695228; rs1940309; rs505277; rs476739; rs867036; rs867035; rs2071674; rs7694946; rs4823324; rs180950; rs2590673; rs892100; rs1792745; rs12546767; rs12634033; rs892270; rs10887149; rs10041417; rs6792971; rs12636438; rs1679255; rs9521789; rs476739; rs505277; rs482700; rs7695228; rs867036; rs867035; rs1940309; rs16889556; rs1047782; rs1530680; rs12667130; rs6965855; rs985080; rs4726782; rs1718101; rs10487689; rs1918296; rs737533; rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; rs12193738; rs2817198; rs10456309; rs985080; rs1554690; rs2533096; rs6951437; rs344470; rs344468; rs4725745; rs12444778; rs1444186; rs2294691; rs10456309; rs1562422; rs807694; rs3756814; rs3777663; rs9295626; rs7763790; rs6935076; rs9348646; rs2328791; rs2328791; rs2817201; rs9295626; rs4576240; rs17307478, rs9356939, rs7763790, rs6456621; rs6456624, rs6935076, rs2038137, rs3756821, rs1883593, rs3212236; rs3777663, rs3756814, rs6931809, rs6916186, rs6933328, rs17491647; rs2328791; rs33914824a; rs807694a; rs707864a; rs10456301a; rs16889066a; rs9379651a; rs2817201; rs9295626; rs10456309; rs4576240; rs17307478; rs9356939; rs7763790; rs6456621; rs3756821; rs1883593; rs3212236; rs2294691;

rs3777663; rs3756814; rs6931809; rs6916186; rs6933328; rs17491647; rs9348646; rs1562422 y marcadores de los siguientes genes: R5H2; OR5H6; RRAGA; OR6B3; UMOD; A26C1A; FAM29A; CHRNA1; IFIT5; LOC643905; K2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; PGD; MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622; OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF.

5 Breve descripción de las figuras

FIG. 1: (A) Estructura del STR BV677278. (B) Ubicación del bloque de haplotipos de riesgo dentro del gen *DCDC2*, en relación con BV677278. Los exones están numerados.

10 FIG. 2: (A-B) resultados de SILAC para células Raji y HeLa; diagrama de interacción bidimensional. (C) Resultados del CHIP. α -H3: anticuerpo contra una variante de histona H3 enriquecida en genes de transcripción activa. β -actina: amplicón de control del gen de la β -actina. Las barras de error representan la desviación estándar entre 3 replicados. El doble asterisco (**) representa un valor de p por debajo de 0,01 (prueba T de una cola; véase la Tabla S6).

15 FIG. 3: (A) Efecto del estado de portador de haplotipos de riesgo en varios fenotipos de la lectura, el lenguaje y la cognición (descritos en la Tabla S1 y en la sección "Materiales y métodos de los ejemplos"). Los puntos de datos representan la media de cada grupo, convertida en una puntuación z relativa a la media de la población de muestra de ALSPAC. Las unidades del eje "y" son fracciones de una desviación estándar. PD: tarea de eliminación de fonemas; Lectura 7: lectura de palabras sueltas a los 7 años de edad; Lectura NW: lectura de pseudopalabras a los 9 años de edad; Ortografía 7 y 9: ortografía a los 7 y 9 años de edad; WOLD: tarea de comprensión verbal de las Dimensiones Objetivas del Aprendizaje de Wechsler; NWR: tarea de repetición de pseudopalabras. (B) Modelo de efectos diferenciales de los alelos de BV677278. Los monómeros de ETV6 deben al menos homodimerizarse a través de sus dominios puntiagudos (PNT) para unirse al ADN a través de sus dominios ETS, y pueden homopolimerizarse *in vivo*. Los indeles de las unidades de repetición de BV677278 podrían cambiar el tamaño del polímero de ETV6 y, por lo tanto, afectar la expresión del gen objetivo.

20 FIG. 4: Árbol filogenético basado en la alineación múltiple de 22 alelos de BV677278. Se utilizó el algoritmo Clustal W, con parámetros de método predeterminados (matriz IUB, penalización por hueco = 15, penalización por extensión de hueco = 6,66). Clado 1, que contiene los alelos de riesgo 5 y 6, es el grupo más ramificado de todos, en gris claro (sin incluir la rama para el alelo 22).

25 FIG. 5: Epistasia de READ1 sobre el haplotipo de riesgo de KIAA0319. (A) Esta gráfica muestra el efecto de tener al menos una copia del alelo READ1 denotado solo (por ejemplo, alelo 5), al menos una copia de KIAHap solo, y al menos una copia de cada uno (ambos), en comparación con todos los miembros de ALSPAC. (B) Esta gráfica muestra los efectos protectores de tener un alelo READ1 que comprende una sola copia de la Unidad de Repetición 1 (RU1_1). Los puntos de datos representan la media transformada en z de cada grupo, en comparación con la media total de ALSPAC (Media de todos), en las medidas indicadas. Las unidades del eje "y" son fracciones de una desviación estándar. Para medir el CI verbal, de rendimiento y total a los 8 años de edad se utilizó la WISC-III. PD: tarea de eliminación de fonemas a los 7 años de edad; Lectura 7: tarea de lectura de palabras sueltas a los 7 años de edad; Lectura 9: tarea de lectura de palabras sueltas a los 9 años de edad; Lectura NW: tarea de lectura de pseudopalabras sueltas a los 9 años de edad, Ortografía 7: tarea de ortografía a los 7 años de edad; Ortografía 9: tarea de ortografía a los 9 años de edad.

30 FIG. 6: (A) Esquema de los genes dentro del locus *DYX2* en el cromosoma 6p21.3. Los genes en gris claro, *DCDC2* y *KIAA0319*, han replicado asociaciones con fenotipos del lenguaje escrito y verbal, a saber, DL y TL. Las regiones en gris oscuro dentro de los genes denotan dos variantes funcionales, READ1 en *DCDC2* y un haplotipo de riesgo con marcadores en *KIAA0319* y *TDP2*, que se han asociado funcionalmente a DL y TL utilizando modelos animales y técnicas moleculares. (B) Un esquema actualizado de genes con marcadores que muestran asociaciones replicadas a DL, TL y/o CI. Los genes (que se muestran en gris claro) se han expandido a siete (*DCDC2*, *KIAA0319*, *TDP2*, *ACOT13*, *C6orf62*, *FAM65B* y *CMAHP*), aunque el desequilibrio del enlace pueda dar cuenta de múltiples asociaciones (en particular para *KIAA0319*, *TDP2*, *ACOT13* y *C6orf62*).

35 Descripción detallada

Las sondas y cebadores polinucleotídicos de una variante de *DCDC2* y una variante de *KIAA0319* que se describen en el presente documento son polinucleótidos recombinantes, aislados y sintéticos que detectan una alteración en un gen *DCDC2* (denominado gen "*DCDC2* variante") que está asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA), tales como polinucleótidos aislados y recombinantes que detectan una alteración de *DCDC2* en el locus *DYX2*. La variante es, por ejemplo, un haplotipo de riesgo de *DCDC2* (por ejemplo, CGCGAG, GACGAG), el alelo 5 de BV677278 (READ1); el alelo 6 de BV677278 (READ1); o una, dos o tres o más de las variantes asociadas a la susceptibilidad a desarrollar una dificultad del aprendizaje (DA). También se describen polinucleótidos aislados, sintéticos y recombinantes que detectan una alteración en un gen *KIAA0319* (denominada "variante del gen *KIAA0319*") que está asociada a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA), tales como polinucleótidos aislados y recombinantes que detectan una alteración de *KIAA0319* en el locus *DYX2*. La variante es, por ejemplo, un haplotipo de riesgo de *KIAA0319* (por ejemplo, rs4504469, rs2038137, rs2143340); o una, dos o tres o más de las variantes asociadas a la susceptibilidad a desarrollar una dificultad del

aprendizaje (DA). La DA es una discapacidad lectora (DL) o un trastorno del lenguaje (TL). Las sondas polinucleotídicas típicamente tienen una secuencia que es total o parcialmente complementaria a la secuencia de la alteración y la región flanqueante, y se hibridan a la alteración de interés y a la secuencia flanqueante de una manera específica. Los polinucleótidos descritos en el presente documento pueden detectar una variedad de alteraciones en un gen *DCDC2* o en un gen *KIAA0319* asociadas a la susceptibilidad de desarrollar DA, tales como DL y TL. Por ejemplo, se puede detectar un polimorfismo de un solo nucleótido (SNP) de una región codificante, exón, límite exón-intrón, péptido señal, región no traducida 5', región promotora, secuencia potenciadora, región no traducida 3' o intrón que está asociado a DA, tales como DL y TL. Estos polimorfismos incluyen, a modo no taxativo, aquellos que dan como resultado cambios en la secuencia de aminoácidos de las proteínas codificadas por el gen *DCDC2*, y cambios en la secuencia de aminoácidos de las proteínas codificadas por el gen *KIAA0319*, generan productos de empalme alternativos, crean productos truncados, introducen un codón de parada prematuro, introducen un exón críptico, alteran el grado o la expresión en mayor o menor grado, alteran la especificidad tisular de *DCDC2* y/o la expresión de *KIAA0319* (por ejemplo, a nivel del ARNm o de la proteína), introducen cambios en la estructura terciaria de las proteínas codificadas por *DCDC2* y/o *KIAA0319*, introducen cambios en la afinidad de unión o especificidad de las proteínas expresadas por *DCDC2* y/o *KIAA0319* o alteran la función de las proteínas codificadas por *DCDC2* y/o *KIAA0319*. Los polinucleótidos en cuestión incluyen polinucleótidos que son variantes de los polinucleótidos descritos en el presente documento, siempre que los polinucleótidos variantes mantengan su capacidad para detectar específicamente una variación en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar DA, como DL y/o TL, o en el gen *KIAA0319* que está asociada a la susceptibilidad para desarrollar DA, tales como DL y/o TL. Los polinucleótidos variantes pueden incluir, por ejemplo, secuencias que difieren en una o más sustituciones, adiciones o deleciones de nucleótidos.

En ciertos aspectos de la divulgación, el polinucleótido aislado o recombinante es una sonda que se hibrida, en condiciones rigurosas, tales como condiciones altamente rigurosas, a una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA, o una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA. Una DA puede ser, por ejemplo, una discapacidad lectora (DL) o trastorno del lenguaje (TL). Tal como se usa en el presente documento, el término "hibridación" se utiliza en referencia al emparejamiento de ácidos nucleicos complementarios. El término "sonda" se refiere a un polinucleótido que es capaz de hibridarse con otro ácido nucleico de interés. El polinucleótido puede ser natural, como en un producto de digestión restringido purificado, o puede producirse de forma sintética, recombinante o mediante amplificación de ácidos nucleicos (por ejemplo, amplificación por PCR).

En la técnica es bien conocido cómo realizar experimentos de hibridación con moléculas de ácidos nucleicos. El experto en la técnica está familiarizado con las condiciones de hibridación y sabe que las condiciones de rigurosidad apropiadas que promueven la hibridación de ADN pueden variar. Estas condiciones de hibridación se mencionan en libros de texto estándar como *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory (1989); y *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley y Sons: 1992. En un aspecto, los polinucleótidos se hibridan con una variación en el gen *DCDC2*, con una variación en un gen *KIAA0319*, o tanto con una variación en el gen *DCDC2* como con una en el gen *KIAA0319* (por ejemplo, el uso de sondas distintas que se hibridan con cada uno de los genes, respectivamente). En condiciones altamente rigurosas, esencialmente no se produce hibridación con polinucleótidos no relacionados.

La hibridación de ácidos nucleicos se ve afectada por condiciones tales como la concentración de sales, la temperatura, los disolventes orgánicos, la composición de la base, la longitud de las cadenas complementarias y el número de emparejamientos erróneos de bases de nucleótidos entre los ácidos nucleicos hibridantes, como sabrán apreciar fácilmente los expertos en la técnica. Las condiciones rigurosas de temperatura generalmente incluirán temperaturas superiores a 30 °C, o pueden ser superiores a 37 °C o 45 °C. Las condiciones rigurosas de sales normalmente serán menores que 1000 mM, o pueden ser menores que 500 mM o 200 mM. Por ejemplo, uno podría realizar la hibridación a 6,0x cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45 °C, seguida de un lavado de 2,0x SSC a 50 °C. Por ejemplo, la concentración de sales en la etapa de lavado se puede seleccionar desde una rigurosidad baja de aproximadamente 2,0x SSC a 50 °C hasta una rigurosidad alta de aproximadamente 0,2x SSC a 50 °C. Además, la temperatura en la etapa de lavado se puede aumentar desde condiciones de baja rigurosidad a temperatura ambiente, aproximadamente 22 °C, hasta condiciones de alta rigurosidad a aproximadamente 65 °C. Tanto la temperatura como la sal pueden variar, o la temperatura o la concentración de sales pueden mantenerse constantes mientras se cambia la otra variable. En un aspecto, los ácidos nucleicos se hibridan en condiciones de baja rigurosidad de 6,0x SSC a temperatura ambiente, seguido de un lavado a 2,0x SSC a temperatura ambiente. Sin embargo, la combinación de parámetros es mucho más importante que la medición de cualquier parámetro individual. Véase, por ejemplo, Wetmur y Davidson, *J Mol Biol.* 1968; 31 (3): 349-70. Las secuencias de sondas también pueden hibridarse específicamente con ADN dúplex, en ciertas condiciones, para formar complejos de ADN tríplex o de orden superior. La preparación de tales sondas y las condiciones de hibridación adecuadas son bien conocidas en la técnica. Un método para obtener ADN que codifica las construcciones biosintéticas divulgadas en el presente documento es mediante el ensamblaje de oligonucleótidos sintéticos producidos en un sintetizador de oligonucleótidos convencional y automatizado.

Una sonda o cebador polinucleotídico utilizado en un método descrito en el presente documento puede marcarse con una molécula reportera, de modo que sea detectable en un sistema de detección, que incluye, entre otros, sistemas enzimáticos (por ejemplo, ELISA, así como ensayos histoquímicos enzimáticos), sistemas fluorescentes,

radioactivos, químicos y luminiscentes. Una sonda o cebador polinucleotídico usado en un método descrito en el presente documento puede incluir además un resto extintor que, cuando se coloca muy cerca de una etiqueta (por ejemplo, una etiqueta fluorescente), hace que se produzca poca o ninguna señal desde la etiqueta. No se pretende que la presente invención se limite a ningún sistema o etiqueta de detección particular.

5 En otro aspecto, el polinucleótido aislado es un cebador que se hibrida de forma adyacente, ascendente o descendente a una alteración en el gen *DCDC2* o el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA en seres humanos. Por ejemplo, un cebador polinucleotídico se puede hibridar de forma adyacente, ascendente o descendente a una alteración en el gen *DCDC2*, o de forma adyacente, ascendente o descendente a una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL). Tal como se usa en el presente documento, el término “cebador” se refiere a un polinucleótido que actúa como un punto de inicio de la síntesis de ácidos nucleicos cuando se coloca en condiciones en las que se induce la síntesis de un producto de extensión de cebador que es complementario a una cadena de ácidos nucleicos (por ejemplo, en la presencia de nucleótidos, un agente inductor como la ADN polimerasa, y la temperatura, el pH y la concentración de electrolitos adecuados). De manera alternativa, el cebador se liga a un ácido nucleico proximal cuando se coloca en condiciones en las que se induce la ligadura de dos ácidos nucleicos no enlazados (por ejemplo, en presencia de un ácido nucleico proximal, un agente inductor como la ADN ligasa, y la temperatura, el pH y la concentración de electrolitos adecuados). Un cebador polinucleotídico puede ser natural, como en un producto de digestión restringido purificado, o puede producirse sintéticamente. El cebador es monocatenario o bicatenario. Si es bicatenario, el cebador se trata para separar sus cadenas antes de ser utilizado. El cebador puede ser un oligodesoxirribonucleótido. Las longitudes exactas de los cebadores dependerán de muchos factores, incluida la temperatura, la fuente del cebador y el método utilizado.

Un aspecto de la divulgación es un par de cebadores que detectan específicamente una alteración en un gen *DCDC2* o una alteración en un gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA. En tal caso, el primer cebador se hibrida corriente arriba de la alteración y un segundo cebador se hibrida corriente abajo de la alteración. Uno de los cebadores se hibrida con una cadena de una región de ADN que comprende una alteración en el gen *DCDC2* o en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar DA, y el segundo cebador se hibrida con la cadena complementaria de una región de ADN que comprende una alteración en el gen *DCDC2* o en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA. Tal como se usa en el presente documento, el término “región de ADN” se refiere a un tramo subcromosómico de ADN. Otros aspectos de la divulgación son pares de cebadores que detectan específicamente alteraciones en otros genes, descritos en el presente documento, que están asociados a una susceptibilidad de desarrollar una dificultad del aprendizaje. Un aspecto adicional de la divulgación es un conjunto de tres cebadores útiles para distinguir entre dos alelos de *DCDC2*, en donde el primer alelo es un gen *DCDC2* no eliminado (por ejemplo, un alelo que no comprende una delección de *READ1*), y el segundo alelo comprende una delección en el gen *DCDC2* (por ejemplo, comprende el alelo 39/indicado “Del” en la Tabla S4, Ejemplo 1) que está asociada a la susceptibilidad de desarrollar DA. El primer cebador se hibrida con una secuencia de nucleótidos que es común a ambos alelos, como una secuencia de nucleótidos no alélica que está corriente arriba o corriente abajo de la secuencia polimórfica en el gen *DCDC2*. El segundo cebador se hibrida específicamente con una secuencia de nucleótidos que es única para un primer alelo (por ejemplo, un gen *DCDC2* no eliminado). El tercer cebador se hibrida específicamente con una secuencia de nucleótidos que es única para el segundo alelo (por ejemplo, una delección en el gen *DCDC2* que está asociada a la susceptibilidad para desarrollar DL). El uso del conjunto de tres cebadores da como resultado la amplificación de una región de ADN que depende de qué alelo de *DCDC2* está presente en la muestra. De manera alternativa, dos cebadores del conjunto se hibridan con una secuencia de nucleótidos que es común a dos alelos del gen *DCDC2*, como las secuencias de nucleótidos no alélicas que están corriente arriba y corriente abajo de una secuencia polimórfica en el gen *DCDC2*, y el tercer cebador se hibrida específicamente a uno de los dos alelos del gen *DCDC2*.

Sondas y cebadores polinucleotídicos de la variante de *DCDC2*

Los polinucleótidos se pueden usar en cualquier ensayo que permita la detección de una variación en el gen *DCDC2* que esté asociada a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL). Dichos métodos pueden abarcar, por ejemplo, métodos de detección mediados por hibridación, mediados por ligación o mediados por extensión de cebador. Además, cualquier combinación de estos métodos puede utilizarse en la invención.

En un aspecto, los polinucleótidos detectan una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA al amplificar una región de ADN que comprende la alteración en el gen. Se puede utilizar cualquier método de amplificación. En un segundo aspecto, los polinucleótidos detectan una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA al amplificar una región de ADN que comprende la alteración en el gen. En un aspecto específico, una región de ADN que comprende una alteración se amplifica utilizando la reacción en cadena de la polimerasa (PCR). (Ann. Rev. Biochem., 61:131-156 (1992)); Gilliland *et al.*, Proc. Natl. Acad. Sci., 87: 2725-2729 (1990); Bevan *et al.*, PCR Methods and Applications, 1: 222-228 (1992); Green *et al.*, PCR Methods and Applications, 1: 77-90 (1991); Blackwell *et al.*, Science, 250: 1104-1110 (1990). “PCR” se refiere, por ejemplo, al método de Mullis (véanse, por ejemplo, las Patentes de los Estados Unidos N.º 4,683,195, 4,683,202 y 4,965,188), que describen un método para aumentar la concentración de una región de ADN, en una mezcla de ADN genómico, sin clonación ni purificación. Por ejemplo, los cebadores polinucleotídicos de la invención, descritos en el presente documento, se combinan con una mezcla de ADN (o cualquier secuencia

de polinucleótidos que pueda amplificarse con los cebadores polinucleotídicos), en donde el ADN comprende el gen *DCDC2* y/o el gen *KIAA0319*. La mezcla también incluye los reactivos de amplificación (por ejemplo, desoxirribonucleótido trifosfatos, amortiguadores, etc.) necesarios para la reacción de ciclado térmico. De acuerdo con los métodos PCR estándar, la mezcla experimenta una serie de pasos de desnaturalización, hibridación del cebador y extensión de la polimerasa para amplificar la región de ADN que comprende una variación en el gen *DCDC2* y/o una variación en el gen *KIAA0319*. La longitud de la región amplificada de ADN está determinada por las posiciones relativas de los cebadores entre sí y, por lo tanto, esta longitud es un parámetro controlable. Por ejemplo, la hibridación de los cebadores puede producirse de tal manera que los extremos de los cebadores proximales a la mutación estén separados por 1 a 10.000 pares de bases (por ejemplo, 10 pares de bases (pb) 50 pb, 200 pb, 500 pb, 1.000 pb, 2.500 pb, 5.000 pb o 10.000 pb).

Se usa instrumentación estándar para la amplificación de ADN y la detección de ADN amplificado. Por ejemplo, se ha desarrollado una amplia variedad de instrumentos para llevar a cabo amplificaciones de ácidos nucleicos, particularmente PCR, por ejemplo, Johnson *et al.*, Patente de los Estados Unidos N.º 5,038,852 (termociclador controlado por ordenador); Wittwer *et al.*, Nucleic Acids Research, 17: 4353-4357 (1989) (PCR de tubos capilares); Hallsby, Patente de los Estados Unidos N.º 5,187,084 (control de temperatura basado en aire); Garner *et al.*, Biotechniques, 14: 112-115 (1993) (PCR de alto rendimiento en placas de 864 pocillos); Wilding *et al.*, Solicitud Internacional N.º PCT/US93/04039 (PCR en estructuras micromecanizadas); Schnipelsky *et al.*, Solicitud de Patente Europea N.º 90301061.9 (publ. N.º 0381501 A2) (dispositivo de PCR desechable de un solo uso). En ciertos aspectos de la divulgación, se usa PCR en tiempo real u otros métodos conocidos en la técnica, tales como el ensayo de Taqman.

El ADN amplificado puede analizarse mediante varios métodos diferentes. Los métodos para analizar el ADN amplificado incluyen, a modo no taxativo, secuenciación (Sanger) del ADN, determinación del tamaño del fragmento por electroforesis o cromatografía, hibridación con una sonda marcada, hibridación con una matriz o micromatriz de ADN, incorporación de cebadores biotinilados seguida de detección de conjugado de avidina-enzima, o incorporación de desoxinucleótidos trifosfatos marcados con ³²P, como dCTP o dATP, en el segmento amplificado. En un aspecto, el ADN amplificado se analiza mediante electroforesis en gel. Los métodos de electroforesis en gel son bien conocidos en la técnica. Véase, por ejemplo, Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley y Sons: 1992. El ADN amplificado se puede visualizar, por ejemplo, por medios fluorescentes o radioactivos. El ADN también puede transferirse a un soporte sólido como una membrana de nitrocelulosa y someterse a transferencia de Southern después de la electroforesis en gel. En un aspecto, el ADN se analiza por electroforesis y se expone a bromuro de etidio, y se visualiza bajo luz ultravioleta.

En un aspecto, la alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar DL es una deleción. La deleción se puede detectar utilizando cebadores polinucleotídicos descritos en el presente documento. Por ejemplo, se puede usar un conjunto de tres cebadores para distinguir entre un alelo del gen *DCDC2* que comprende una deleción y un gen *DCDC2* de tipo natural. El uso del conjunto de tres cebadores da como resultado la amplificación de una región de ADN que depende de qué alelo de *DCDC2* está presente en la muestra. En algunos casos, una deleción es protectora, como el alelo 39/Del en la Tabla S4 del Ejemplo 1. En algunos aspectos de la divulgación, las modificaciones o variantes son protectoras.

En otro aspecto, el ADN amplificado se analiza mediante secuenciación de ADN. La determinación de la secuencia de ADN se puede realizar mediante métodos estándar como la tecnología de terminación de cadena dideoxi (secuenciación de Sanger) y electroforesis en gel, o mediante otros métodos como la pirosecuenciación (Biotage AB, Uppsala, Suecia). La secuencia de ácidos nucleicos del ADN amplificado se puede comparar con la secuencia de ácidos nucleicos del ADN de tipo natural para identificar si existe una variación en el gen *DCDC2* y/o *KIAA0319* que está asociada a la susceptibilidad de desarrollar DA.

En otro aspecto, los polinucleótidos de la divulgación detectan una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA mediante métodos mediados por hibridación. En un aspecto adicional, los polinucleótidos detectan una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA mediante métodos mediados por hibridación. En un aspecto, una sonda polinucleotídica se hibrida con una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA (y con nucleótidos flanqueantes), pero no con un gen *DCDC2* de tipo natural. En otro aspecto, una sonda polinucleotídica se hibrida con una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar un DA (y nucleótidos flanqueantes), pero no con un gen *KIAA0319* de tipo natural. La sonda polinucleotídica puede comprender nucleótidos que están marcados de manera fluorescente, radioactiva o química para facilitar la detección de la hibridación. La hibridación puede realizarse y detectarse mediante métodos estándar conocidos en la técnica como, por ejemplo, mediante transferencia de Northern, transferencia de Southern, hibridación *in situ* fluorescente (FISH) o hibridación con polinucleótidos en un soporte sólido (por ejemplo, matrices, micromatrices de ADN, matrices de ADNc o chips Affymetrix). En un aspecto, la sonda polinucleotídica se usa para hibridar el ADN genómico mediante FISH. El método FISH se puede usar, por ejemplo, en células en metafase, para detectar una deleción en el ADN genómico. Por medio de FISH, el ADN genómico se desnaturaliza para separar las cadenas complementarias dentro de la estructura de doble hélice del ADN. La sonda polinucleotídica se combina con el ADN genómico desnaturalizado. Si está presente una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA o una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una

DA, la sonda se hibridará al ADN genómico. La señal de la sonda (por ejemplo, fluorescencia) se puede detectar a través de un microscopio fluorescente para detectar la presencia o ausencia de señal. La ausencia de señal indica la ausencia de una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL) o la ausencia de una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL). De manera alternativa, la presencia de señal puede usarse para determinar la ausencia de una alteración en el gen *DCDC2* o *KIAA0319*.

En otro aspecto, los polinucleótidos detectan una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar un DA (por ejemplo, DL, TL) o una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA (por ejemplo, DL, TL) por extensión de cebador con ADN polimerasa. En un aspecto, un cebador polinucleotídico se hibrida de forma inmediatamente adyacente a la alteración. Se puede usar una reacción de secuenciación de una sola base utilizando terminadores de didesoxinucleótidos marcados para detectar la alteración. Si hay una alteración presente, el terminador marcado se incorporará al producto de extensión; si no hay una alteración presente, el terminador marcado no se incorporará. En otro aspecto, un cebador polinucleotídico se hibrida a una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA o una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA. El cebador, o una porción de este, no se hibridará con un gen *DCDC2* de tipo natural o un gen *KIAA0319* de tipo natural. Si hay una alteración presente, se produce la extensión del cebador; si no hay una alteración presente, no se produce la extensión del cebador. Los cebadores y/o nucleótidos pueden incluir además sondas fluorescentes, radioactivas o químicas. Un cebador marcado por la extensión del cebador puede detectarse midiendo la intensidad del producto de extensión, como por ejemplo mediante electroforesis en gel, espectrometría de masas o cualquier otro método para detectar marcas fluorescentes, radioactivas o químicas.

En otro aspecto, los polinucleótidos detectan una alteración en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA o una alteración en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA mediante ligación. En un aspecto, un cebador polinucleotídico se hibrida con una variación en el gen *DCDC2* que está asociada a la susceptibilidad de desarrollar una DA o con una variación en el gen *KIAA0319* que está asociada a la susceptibilidad de desarrollar una DA. El cebador no se hibridará con el gen de tipo natural (por ejemplo, el gen *DCDC2* de tipo natural). También se describe un segundo polinucleótido que se hibrida con una región del gen *DCDC2* inmediatamente adyacente al primer cebador o con una región del gen *KIAA0319* inmediatamente adyacente al primer cebador. Uno, o ambos, de los cebadores polinucleotídicos pueden estar marcados por fluorescencia, radioactividad o químicamente. La ligación de los dos cebadores polinucleotídicos se produce en presencia de ADN ligasa si existe una alteración en el gen (por ejemplo, una alteración del gen *DCDC2* o *KIAA0319*) que está asociada a la susceptibilidad de desarrollar una DA. La ligación puede detectarse mediante electroforesis en gel, espectrometría de masas o mediante la medición de la intensidad de marcadores fluorescentes, radioactivos o químicos.

Ejemplos

Los siguientes ejemplos tienen fines ilustrativos y no pretenden ser limitativos de ninguna manera.

Ejemplo 1

Materiales y métodos

Reclutamiento de sujetos, recopilación de datos y ADN, y gestión de datos.

El equipo de ALSPAC realizó el reclutamiento de sujetos y la recopilación de datos de fenotipo y ADN para la cohorte ALSPAC, bajo la supervisión de S.M. Ring, y estos datos fueron manejados para este estudio por L.L. Miller.

El Estudio Longitudinal Avon de Padres e Hijos (ALSPAC)

El ALSPAC es una cohorte de nacimientos prospectiva con sede en la región de Avon del Reino Unido. Se compone de niños, en su mayoría de ascendencia del norte de Europa, nacidos en 1991 y 1992. Los niños fueron reclutados antes del nacimiento; el reclutamiento de sus madres embarazadas dio lugar a un total de 15.458 fetos, de los cuales 14.701 estaban vivos al año de edad. Los detalles sobre los participantes, el reclutamiento y las metodologías de estudio se describen en detalle en otra parte (bristol.ac.uk/alspac). Los niños de ALSPAC han sido ampliamente fenotipados desde antes del nacimiento hasta la edad adulta temprana. Una actualización sobre el estado de la cohorte se publicó en 2012. (S11). Las mediciones de lectura, del lenguaje y cognitivas utilizadas para este estudio se recopilaban a los 7, 8 y 9 años de edad. Las muestras de ADN de 10.676 de estos niños estaban disponibles para el genotipado. De este subconjunto, la cantidad de niños que completaron las mediciones del lenguaje y cognitivas varía según la medición, pero generalmente es de 5200-5600 sujetos.

Mediciones de lectura de ALSPAC

Las mediciones de lectura en ALSPAC incluyen una tarea de eliminación de fonemas a los 7 años de edad, la lectura de palabras sueltas a los 7 y 9 años de edad, la ortografía a los 7 y 9 años de edad, la lectura de seudopalabras sueltas a los 9 años de edad, y la comprensión de un texto, velocidad y precisión a los 9 años de

edad. La tarea de eliminación de fonemas mide el reconocimiento de los fonemas (S12), que se considera en general como un déficit central en la DL (S13). El niño escucha una palabra hablada en voz alta, y luego se le pide que elimine un fonema específico de esa palabra para formar una nueva palabra (por ejemplo, ¿qué palabra se crea cuando el sonido /b/ se elimina de la palabra “block” (bloqueo)? “Lock” (cerradura)). Esta tarea también se conoce como Prueba de Análisis Auditivo y fue desarrollada por Rosner y Simon (S14). La lectura de palabras sueltas se evaluó a los 7 años de edad utilizando la subprueba de lectura de las Dimensiones Objetivas de la Lectura de Wechsler (WORD) (S15). A los 7 y 9 años de edad, se evaluó la ortografía; se le pidió al niño que escribiera un conjunto de 15 palabras adecuadas a su edad (S15). A los 9 años de edad, se volvió a evaluar la lectura de palabras sueltas pidiéndole al niño que leyera diez palabras reales y diez seudopalabras en voz alta. Las palabras y seudopalabras utilizadas son un subconjunto de una lista más grande de palabras y seudopalabras tomadas de la investigación realizada por Terezinha Nunes y otros en Oxford (S16). Las puntuaciones de velocidad, precisión y comprensión lectoras se determinaron a los 9 años de edad, utilizando el Análisis de la Capacidad Lectora de Neale (NARA-11) (S17). Las tres mediciones están estandarizadas. El niño leyó pasajes de un folleto en voz alta e, inmediatamente después, se le hicieron preguntas acerca de lo que había leído para evaluar su comprensión lectora. La precisión se midió contando el número de errores (errores de pronunciación, sustituciones, etc.) que cometió el niño y convirtiéndolo en una puntuación estandarizada. La velocidad de lectura era el número de palabras por minuto.

Mediciones del lenguaje de ALSPAC

Las mediciones del lenguaje en las que se centra este estudio se determinaron a los 8 años de edad. La primera de ellas es una tarea de repetición de seudopalabras (NWR), en la que se le pide al niño que repita las seudopalabras grabadas. Esta tarea mide la memoria fonológica a corto plazo y el procesamiento (S18). La segunda es una subprueba de las Dimensiones Objetivas del Lenguaje de Wechsler que mide la comprensión del lenguaje (WOLD Comp) (S19). Para esta tarea, al niño se le hace una serie de preguntas sobre un párrafo que describe una imagen, que fue leído en voz alta por un examinador. Los niños con TL tienen constantemente un desempeño deficiente en estas mediciones (S20, S21).

Evaluación del CI de ALSPAC

El CI verbal, de rendimiento y total se evaluaron a los 8 años de edad, utilizando la Escala de Inteligencia de Wechsler para Niños (WISC-III).

Aprobación ética

La aprobación ética para el estudio se obtuvo del Comité de Ética y Ley de ALSPAC, el Comité de Ética de Investigación local del Reino Unido y el Comité de Investigación con Seres Humanos de Yale.

Diseño y genotipado de paneles TagSNP en DYX2.

Los TagSNP diseñados para capturar la variación común en el locus DYX2 se seleccionaron usando el servidor de diseño de estudios de asociaciones de Han *et al.* (design.cs.ucla.edu) (33). Los SNP fueron genotipados en la plataforma Sequenom, en colaboración con el Centro de Análisis del Genoma de Yale (West Haven, CT), según protocolos estándar. La tasa de llamadas y las estadísticas descriptivas de los SNP descritos en el presente documento se enumeran en la Tabla S4. rs4504469, rs2038137 y rs2143340 fueron genotipados por Scerri *et al.*, como se describe (24). La frecuencia de alelos menores para todos los tagSNP fue mayor o igual a 0,05. La potencia promedio para capturar variantes comunes conocidas (FAM >0,05) dentro de DYX2 usando este panel se estimó en 83,44 % *a priori*. Se incluyeron varios otros SNP además del panel tagSNP, incluidos varios que anteriormente se habían asociado con DL y SNP codificaciones en DCDC2.

Análisis de asociaciones basado en haplotipos.

Se evaluó el desequilibrio de enlace y se definieron los haplotipos utilizando el paquete de software Haploview, versión 4.2 (34). Los marcadores que se desviaban sustancialmente del equilibrio de Hardy-Weinberg, o que tenían una tasa de llamada <85 %, no se utilizaron para el análisis de haplotipos. Para demarcar los bloques de haplotipos se usó la opción de la regla de los cuatro gametos, lo que dio como resultado 44 bloques de haplotipos que cubren el locus DYX2. El análisis de asociaciones se realizó con haplotipos individuales que tenían frecuencias de 0,01 o más (208 en total), utilizando el paquete de software Plink, versión 1.07 (S4). Los análisis de asociaciones se realizaron utilizando estadísticas de prueba de regresión logística y chi cuadrado (opciones -hap-assoc y -hap-logistic). Los individuos que no fueron identificados como blancos no hispanos, aquellos que tenían un coeficiente intelectual total inferior a 75 o cuya muestra de ADN devolvió una tasa de llamada promedio inferior al 85 % para los SNP que pasaron el control de calidad, se excluyeron del análisis de asociaciones. Para corregir las pruebas múltiples, se aplicó una corrección de Bonferroni con el nivel alfa establecido en 0,05, tratando a cada uno de los 208 haplotipos como una prueba individual; por lo tanto, el nivel de umbral es $0,05/208 = 2,4038 \times 10^{-4}$. Como cada fenotipo constituía una hipótesis independiente (reconocimiento fonológico y lenguaje), este umbral no se duplicó para tener en cuenta que había dos fenotipos.

Genotipado de BV677278.

ES 2 718 343 T3

Los portadores de los haplotipos de DCDC2 de interés que se podían eliminar de forma inequívoca (utilizando la función -hap-phase de Plink) se genotiparon para el STR BV677278. BV677278 está genotipado por amplificación por PCR y subsiguiente secuenciación de Sanger. Los alelos se llaman por un script interno de Perl desarrollado por Y. Kong. El script de Perl está disponible a petición.

5 *Cebadores de amplificación*

STR_F: 5'-TGTA AACGACGGCCAGTTGTTGAATCCCAGACCACAA-3' (SEQ ID NO:1)

STR_R: 5'-ATCCCGATGAAATGAAAAGG-3' (SEQ ID NO:2)

Cebador de secuenciación M13F

5'-TGTA AACGACGGCCAGT (SEQ ID NO:3)

10 *Mezcla de reacción de amplificación (por reacción de 10 µl)*

Amortiguador para PCR 10X (Qiagen): 1 µl

MgCl₂ (25 mM) (Qiagen): 0,4 µl

dNTP (10 mM): 0,25 µl

Cebador STR_F (10 µM): 0,25 µl

15 Cebador STR_R (10 µM): 0,25 µl

HotStarTaq™ (Qiagen, diluido a 0,5 unidades/µl en amortiguador de dilución Taq): 0,20 µl

ADN de plantilla: 1 µl (~10 ng/µl)

H₂O sin nucleasa: 6,65 µl

Reacción de amplificación

20 1. 15 minutos, 95 °C

2. 30 segundos, 95 °C

3. 30 segundos, 65 °C

-Disminuir 1 °C/ciclo

4. 60 segundos, 72 °C

25 5. Ir al paso 2, 9 veces

6. 30 segundos, 95 °C

7. 30 segundos, 56 °C

8. 60 segundos, 72 °C

9. Ir al paso 6, 34 veces

30 10. 5 minutos, 72 °C

11. ∞, 4 °C

Purificación y secuenciación de PCR

Los productos de PCR se purificaron usando la mezcla de enzimas ExoSAP-IT®, de acuerdo con el protocolo del fabricante. Los amplicones purificados se mezclaron luego con el cebador de secuenciación M13F y se secuenciaron. La secuenciación de Sanger se realizó en el Centro de Secuenciación de ADN W.M. Keck de Yale, según su protocolo de secuenciación estándar.

35

Tipificación de genotipos

Los alelos fueron tipificados desde los electroferogramas, usando un script de Perl interno desarrollado por Y. Kong para este propósito.

40 Genotipado de microdelección

Los portadores de los haplotipos de *DCDC2* de interés también fueron genotipados para la microdelección de 2.445 pb de *DCDC2*. Esta eliminación abarca todo el STR BV677278 dentro de sus puntos de ruptura, por lo que debe ser genotipado además de BV677278 para obtener un genotipo preciso para homocigotos BV677278 aparentes. La microdelección se genotipa mediante PCR específica de alelos y electroforesis en agarosa. La reacción de tres cebadores genera un amplicón de ~600 pb a partir de cromosomas intactos (sin deleción), y un amplicón de ~200 pb de cromosomas con la deleción, lo que permite que los heterocigotos y ambos homocigotos sean fácilmente distinguibles entre sí.

Cebadores de amplificación

Cebador Del_F: 5'-AGCCTGCCTACCACAGAGAA-3' (SEQ ID NO:4)

10 Cebador Del_RC: 5'-GGAACAACCTCACAGAAATGG-3' (SEQ ID NO:5)

Cebador Del_RD: 5'-TGAAACCCCGTCTCTACTGAA-3' (SEQ ID NO:6)

Mezcla de reacción de amplificación (por reacción de 10 µl)

Amortiguador para PCR 10X (Qiagen): 1 µl

MgCl₂ (25 mM) (Qiagen): 0,4 µl

15 dNTP (10 mM): 0,25 µl

Cebador Del_F (10 µM): 0,30 µl

Cebador Del_RC (10 µM): 0,20 µl

Cebador Del_RD: 0,20 µl

HotStarTaq™ (Qiagen, diluido a 0,5 unidades/µl en amortiguador de dilución Taq): 0,20 µl

20 ADN de plantilla: 1 µl (~10 ng/µl)

H₂O sin nucleasa: 6,45 µl

Reacción de amplificación

La reacción de amplificación para la microdelección es la misma que para el STR BV677278 (véase arriba).

Electroforesis en agarosa

25 Los productos de PCR se sometieron a electroforesis en geles de agarosa al 1 %, utilizando un amortiguador TBE 1X estándar con bromuro de etidio (0,2 µg/ml), a través de métodos estándar, a 100-150 V, dependiendo del tamaño del gel. Se tomaron imágenes de los geles en un transiluminador UV y se documentaron con un sistema de imágenes XR Bio-Rad Gel Doc™. Los genotipos fueron obtenidos de los geles manualmente.

Identificación de proteínas por espectrometría de masas basada en SILAC.

30 Las células Raji y HeLa se marcaron con SILAC con Lys-8 y Arg-10 (Eurisotop) o sus contrapartes naturales Lys-0, Arg-10 (Sigma), como se describe (20). El lisado nuclear pesado preparado a partir de estas células se incubó con una sonda oligonucleotídica biotinilada idéntica a un segmento de BV677278 que se había demostrado previamente que se unía a una proteína nuclear con alta especificidad (15). El lisado nuclear ligero se incubó con una sonda encriptada biotinilada que previamente no se había demostrado que se unía a la proteína nuclear de interés (15).
35 Los complejos de oligonucleótidos y proteínas obtenidos se eliminaron con perlas conjugadas con estreptavidina y se sometieron a espectrometría de masas cuantitativa, como se describió anteriormente (36). También se realizó el experimento inverso (sonda de unión con lisado ligero, sonda encriptada con lisado pesado), dando como resultado los diagramas de interacción bidimensional de la Fig. 3A-B. El experimento anterior se describe con más detalle a continuación:

40 *Marcado con SILAC de células HeLa y Raji*

Se marcaron células Raji para al menos 8 generaciones en medio DMEM (-Arg, -Lys) que contenía suero bovino fetal dializado al 10 % (Gibco) suplementado con 58 mg/L de L-arginina 13C615N4 y 34 mg/L de L- lisina 13C615N2 (Eurisotop), o los correspondientes aminoácidos no marcados. Para Raji, los extractos celulares se prepararon como se describe en Wu *et al.* (S5). Las células HeLa S3 se marcaron con SILAC en medio RPMI 1640 (-Arg, -Lys) que contenía suero bovino fetal dializado al 10 % (Gibco) suplementado con 84 mg/L de L-arginina 13C615N4 y 40 mg/L de L-lisina 13C615N2 (Eurisotop), o los correspondientes aminoácidos no marcados, respectivamente. Para HeLa S3, tres lotes consecutivos de células se recolectaron independientemente, y los extractos celulares se prepararon como describe por Dignam *et al.* (S6).

Se realizaron SILAC, extracción de ADN de proteínas y espectrometría de masas cuantitativa como se describió previamente (S7), utilizando las líneas celulares de Raji y HeLa. La sonda desplegable de unión es un concatémero de dos copias de la sonda EMSA3 utilizada en los experimentos EMSA que informamos en 2011, mientras que la sonda encriptada es un concatémero de dos copias de la sonda EMSA3-Scram1 de los mismos experimentos (S8).

Oligonucleótidos:

EMSA3_for:

5'-TTGAGAGGAAGGAAAGGAAGGATCCCTGAGAGGAAGGAAAGGAAGGA-3' (SEQ ID NO:7)

EMSA3_rev:

10 5'-AATCCTTCCTTCCTTCCTCAGGGA TCCTTCCTTCCTTCCTC-3' (SEQ ID NO:8)

EMSA3/Scram1_for:

5'-TTGAGAGAGAGAGAGAGAGATCCCTGAGAGAGAGAGAGAGAGAGA-3' (SEQ ID NO:9)

EMSA3/Scram1_rev:

5'-AATCTCTCTCTCTCTCTCAGGGATCTCTCTCTCTCTCTCTC-3' (SEQ ID NO:10)

15 *Ensayo desplegable de ADN*

Se unieron 25 µg de sondas de ADN hibridadas, concatenadas y destiobiotiniladas a 75 µl de Dynabeads MyOne C1 (Life Technologies). El exceso de oligonucleótidos se eliminó y las perlas se incubaron con 400 µg de extractos nucleares marcados con SILAC en amortiguador de unión a proteínas (NaCl 150 mM, Tris-HCl 50 mM (pH 8), NP-40 al 0,5 %, MgCl₂ 10 mM, cóctel inhibidor de proteasa; Roche). Después de 1 h en una rueda de rotación a 4 °C, las perlas se lavaron tres veces, se combinaron, y los complejos de ADN-proteína se eluyeron en un amortiguador de unión a proteínas que contenía biotina 16 mM. El sobrenadante se precipitó con 4 volúmenes (v/v) de etanol durante la noche y las proteínas se sedimentaron por centrifugación máxima en una microcentrífuga de mesa. El sedimento se resolubilizó en urea 8 M/Tris 50 mM, pH 8,0, se redujo con DTT 1 mM, se alquiló con yodoacetamida 3 mM y, posteriormente, se digirió con tripsina (Promega) en amortiguador de bicarbonato de amonio 50 mM, pH 8, a temperatura ambiente durante toda la noche. Las muestras se almacenaron en las puntas de pipeta y se eluyeron antes de su uso.

Espectrometría de masas

Los péptidos se separaron con un gradiente de 140 minutos de acetonitrilo al 5 a 60 por ciento (EasyHPLC, Thermo Fisher) utilizando un capilar de 15 cm, 75 µm, lleno con 3,0 µm de perlas C18 (Dr. Maisch) montado directamente en un espectrómetro de masas LTQ-Orbitrap (Thermo Fisher). El instrumento fue operado en modo de adquisición top10 dependiente de datos. Los datos sin procesar se buscaron utilizando el paquete de software MaxQuant (versión 1.2.0.18) con la base de datos humana completa de IPI (v3.68, 87061 entradas). La especificidad de búsqueda de enzimas fue tripsina/p, con 2 errores permitidos. La carbamidometilación se estableció como modificación fija, mientras que la oxidación de metionina y la N-acetilación de proteína se consideraron modificaciones variables. La búsqueda se realizó con una tolerancia de masa inicial de 7 ppm de precisión de masa para el ion precursor y 0,5 Da para los espectros de MS/MS.

ChIP-QPCR.

Se usó el kit AbCam ChIP (cat. # Ab500) para realizar los ensayos de ChIP descritos, de acuerdo con las instrucciones del fabricante, pero con varias modificaciones que se describen a continuación. Para qPCR, se utilizó el kit qPCR Qiagen QuantiTect® SYBR® Green para ChIP-qPCR. Se siguieron las instrucciones del fabricante de 25 pg de plantilla por reacción. Todas las reacciones por triplicado. Los anticuerpos, las secuencias de cebadores y los métodos detallados para este experimento están disponibles en la "Información de respaldo". Los datos de control de calidad para qPCR se muestran en la Fig. S3.

Paso 1: Fijación y recolección celular

- 45 • Para cada conjunto de 3 reacciones de ChIP, se usaron 9 millones de células recién cosechadas (Raji) en lugar de 3 millones, con volúmenes apropiados para 9 millones de células, como se indica en el protocolo.

Paso 2: Lisis celular

- 50 • Después del lavado final en PBS en el paso de fijación celular, cada parte de 9 millones de células se llevó al paso de lisis celular con volúmenes apropiados para 3 millones de células (9 millones se trataron como si fueran 3 millones, dando como resultado 3 millones de células por reacción de ChIP, en lugar de 1 millón).
- Se reemplazaron 100 µl de amortiguador D con inhibidores de proteasa por ~150 µl de lo siguiente:

Amortiguador D2/PI (S9)

ES 2 718 343 T3

Tris 10 mM

EDTA 1 mM

EGTA 0,5 mM

pH = 7

5 Cóctel inhibidor de la proteasa 1X

- Se usó un sonicador de sonda Branson 450 para la sonicación. El ADN se fragmentó con 18 pulsos de 20 segundos, con la amplitud establecida en 6 y el DC establecido en continuo. Las muestras se mantuvieron en agua helada durante la sonicación y se dejaron reposar 2 minutos en hielo entre cada pulso.

Paso 3: Inmunoprecipitación

10 • Se siguió el protocolo del fabricante.

◦ Anticuerpo α -ETV6: sc-166835X (Santa Cruz Biotech), 5 μ g/ChIP

◦ Anticuerpo de control: ab1791 (AbCam) (histona H3 variante α , enriquecida en genes de transcripción activa, una parte de esto viene con el kit), 2 μ g/ChIP

15 • En lugar de las perlas provistas en el kit, se usaron perlas magnéticas Magna-Bind™ Protein A/G y se usó un soporte magnético, de acuerdo con el protocolo del fabricante, en lugar de la centrifuga, para la inmunoprecipitación.

Etapas 4: Purificación de ADN

20 • Después de revertir los enlaces cruzados y el tratamiento con proteinasa K, el ADN se purificó utilizando un kit de purificación por PCR Qiagen QIAquick® en lugar de la suspensión purificadora de ADN proporcionada en el kit de ChIP.

- Después de la purificación con una columna QIAquick®, cada producto final se eluyó con un total de 60 μ l de amortiguador EB, en dos etapas de elución de 30 μ l. El producto final se cuantificó por fluorescencia (kit de ensayo Quant-IT™ PicoGreen® dsDNA, según el protocolo del fabricante).

qPCR

25 • Se usó el kit qPCR Qiagen QuantiTect® SYBR® Green para ChIP-qPCR, siguiendo las instrucciones del fabricante.

- Se usó una plantilla de 25 pg por reacción.

- Se usaron 1,25 μ l de cada cebador (5 μ M) por reacción.

30 • Se realizó una curva estándar de log-5 utilizando el ADN de Raji de entrada (sonicado, pero no sometido a ChIP). Véase la Fig. S3A-B.

- Para evitar el error de pipeteo, todas las muestras se diluyeron a la misma concentración que la muestra menos concentrada, por lo que se podría agregar la misma cantidad a cada reacción. Las diluciones se realizaron en serie.

- Se realizaron todas las reacciones por triplicado.

35 *Cebadores qPCR*

ChIP-STR (amplicón BV677278)

Cebador ChIPSTR_F: 5'-TCATGCAAAGTTCCAAAACC-3' (SEQ ID NO:11)

Cebador ChIPSTR_R: 5'-GATTCCTCCCTCCCTTCC-3' (SEQ ID NO:12)

Estos cebadores capturan toda la repetición de BV677278 y generan un amplicón de ~200 pb.

40 β -actina (+ control)

Cebador β Act_F: 5'-GCCCTAGGCACCAGGGTGTGA-3' (SEQ ID NO:13)

Cebador β Act_R: 5'-ACAGGGTGCTCCTCAGGGGC-3' (SEQ ID NO:14)

Estos cebadores amplifican una secuencia de ~150 pb del gen de la β -actina que se transcribe activamente.

Reacción de qPCR

1. 15 minutos, 95 °C
- 5 2. 30 segundos, 95 °C
3. 30 segundos, 65 °C
- Disminuir 1 °C por ciclo
4. 60 segundos, 72 °C
5. Lectura de placa
- 10 6. Ir al paso 2, 9 veces
7. 30 segundos, 95 °C
8. 30 segundos, 56 °C
9. 60 segundos, 72 °C
10. Lectura de placa
- 15 11. Ir al paso 6, 39 veces
12. 5 minutos, 72 °C
13. Curva de fusión
14. ∞ , 4 °C

*En la Figura S3, se muestran los datos de control de calidad para el experimento ChIP mostrado aquí.

20 *Cálculos de enriquecimiento*

El enriquecimiento multiplicado se calculó con respecto al control sin anticuerpos (una reacción completa de ChIP, pero sin perlas de solo anticuerpo). Brevemente, esto se hizo aumentando 2 a la potencia negativa de la diferencia entre la C(t) de una condición experimental y su respectivo control sin anticuerpos: Enriquecimiento multiplicado = $2^{-(C(t)_{Exp} - C(t)_{Sin\ anticuerpo})}$

25 Resultados

Dos haplotipos de seis marcadores en DCDC2 están asociados a un rendimiento reducido en las mediciones de lectura y del lenguaje

Para determinar si DCDC2, KIAA0319, ambos o ninguno de los dos genes es responsable de la señal en DYX2, se diseñó un panel tagSNP para cubrir densamente el locus DYX2. Se realizó un análisis de asociación basado en haplotipos de la lectura y el lenguaje en una gran cohorte de nacimiento fenotípicamente extensa: el Estudio Longitudinal Avon de Padres e Hijos (ALSPAC) (17). El análisis mostró un bloque de haplotipos de seis marcadores dentro de DCDC2, de los cuales dos haplotipos (CGCGAG y GACGAG) se asociaron a un rendimiento muy bajo en una tarea de eliminación de fonemas y una medición compuesta del lenguaje, respectivamente (Tabla 1). Para este análisis, los casos de DL se definieron como individuos con una puntuación de dos o más desviaciones estándar por debajo de la media en la tarea de eliminación de fonemas, y los casos de TL como individuos con una puntuación de dos o más desviaciones estándar por debajo de la media en cualquiera de las dos medidas del lenguaje (WOLD/NWR). La tarea de eliminación de fonemas mide el reconocimiento fonológico, que se considera en general como el déficit principal en la DL (2). Las tareas de comprensión verbal y de repetición de pseudopalabras (NWR) de las Dimensiones Objetivas del Aprendizaje de Wechsler (WOLD) que comprenden la medición compuesta del lenguaje WOLD/NWR se utilizan para evaluar las habilidades lingüísticas deficientes; los niños con TL muestran un rendimiento sistemáticamente bajo en estas mediciones (18, 19) (véase la Tabla S1 y los "Materiales y métodos" para obtener más información sobre estas mediciones fenotípicas). Los casos se definieron de esta manera para examinar la asociación de variantes de DYX2 con DL y TL graves. Los dos haplotipos muestran una fuerte asociación con sus respectivos fenotipos; la asociación CGCGAG-DL sobrevivió a la corrección de Bonferroni para pruebas múltiples y el valor p de GACGAG-TL estaba justo por debajo del umbral. Sin embargo, las asociaciones por sí mismas no son lo suficientemente fuertes como para descartar un error de tipo I, en parte debido a las bajas frecuencias de los haplotipos y al pequeño número de casos. Curiosamente, sin embargo, el efecto de estos

haplotipos es lo suficientemente fuerte como para reducir el rendimiento medio en medidas fenotípicas relevantes en portadores versus no portadores. Los portadores del haplotipo CGCGAG, en promedio, mostraron un rendimiento significativamente peor en ocho mediciones relacionadas con la lectura en comparación con los no portadores. Del mismo modo, los portadores del haplotipo GACGAG mostraron un rendimiento promedio significativamente más bajo en la medición compuesta del lenguaje WOLD/NWR (Tabla 2). Este efecto cuantitativo indicó que este hallazgo no es un falso positivo y provocó un análisis adicional. Además, este bloque de haplotipos se encuentra muy cerca de BV677278, un compuesto supuestamente funcional de repetición corta en tándem (STR) del que los solicitantes ya informamos con anterioridad (11) (Fig. 1C). El polimorfismo de BV677278 se deriva de cinco unidades de repetición discretas que varían en número (Fig. 1A, Tabla S4). Esta STR evoluciona rápidamente, como lo indica su alto grado de polimorfismo entre las especies de primates y dentro del *Homo sapiens* (Fig. 1B, Fig. S2, Tabla S4).

Los haplotipos de riesgo de DCDC2 están en un fuerte desequilibrio de enlace con dos alelos de BV677278

El bloque de haplotipos asociado es adyacente a BV677278 (Fig. 1C) y se evaluó si los dos haplotipos de riesgo podrían capturar la asociación que surge de los alelos funcionales de BV677278 a través del desequilibrio de enlace. Para abordar esta pregunta, todos los portadores de estos haplotipos fueron sometidos a genotipado de BV677278 por secuenciación de Sanger. De los portadores del haplotipo CGCGAG, el 92 % también tiene el alelo 5 de BV677278. Asimismo, el 78 % de los portadores de GACGAG también portan el alelo 6 de BV677278 (Tabla 1). Los alelos 5 y 6 son similares en estructura, y se agrupan filogenéticamente al mismo clado (Tabla S4, Fig. S1). De hecho, casi todos los portadores de estos dos haplotipos también portan un alelo de este clado (Tabla 1). Estos resultados implican aún más a BV677278 como una variante de riesgo de DL y lo amplían como una posible variante de riesgo de TL (11), y junto con su capacidad reguladora aparente, sugieren que estos alelos de BV677278 son responsables de los efectos de los haplotipos de riesgo.

BV677278 se une específicamente al factor de transcripción ETV6

Para obtener información mecánica sobre la función de BV677278, se usó espectrometría de masas cuantitativa para identificar la(s) proteína(s) que se unen a este locus (20). Con este fin, una sonda oligonucleotídica biotinilada que llevaba segmentos de la repetición de BV677278, que previamente se había demostrado que se unía a una proteína nuclear, y un control no enlazado encriptado, se incubaron con extractos nucleares que habían sido marcados con SILAC (15). El marcado de SILAC implica cultivar dos poblaciones paralelas de células: una con medios que contienen aminoácidos marcados con isótopos pesados de carbono y nitrógeno, y la otra con isótopos naturales. Una vez que se incorpora la marca, las proteínas de las dos poblaciones ("pesada" y "ligera") se pueden diferenciar entre sí mediante espectrometría de masas cuantitativa. El extracto nuclear pesado se incubó con la sonda de BV677278 y el extracto nuclear ligero con la sonda de control. Después, las sondas se extrajeron con perlas conjugadas con estreptavidina y la mezcla de proteínas obtenida se sometió a espectrometría de masas cuantitativa, y se buscaron proteínas que se enriquecieron significativamente con la sonda de BV677278 en comparación con la sonda de control (alto ratio pesada: ligera). El experimento se realizó con extractos nucleares derivados de células HeLa o Raji, y se repitió con una conmutación de marcas que dio como resultado un diagrama de interacción bidimensional. Estos experimentos produjeron un único candidato, compartido tanto por HeLa como por Raji: el factor de transcripción ETV6 (Fig. 2A-B). Para confirmar la interacción BV677278-ETV6, se realizó una inmunoprecipitación de cromatina con PCR cuantitativa (ChIP-qPCR) utilizando anticuerpo α -ETV6, en células Raji. La inmunoprecipitación con el anticuerpo α -ETV6 mostró un pronunciado enriquecimiento para el amplicón de BV677278, pero no para el amplicón de control derivado del gen que codifica la β -actina (*ACTB*) (Fig. 2C). Estos resultados demostraron que ETV6 se une a la región de BV677278.

Los haplotipos de riesgo de DCDC2 muestran una interacción genética sinérgica con un haplotipo de riesgo de DL conocido en el gen KIAA0319

Junto con los hallazgos anteriores del solicitante, estos datos implicaron a BV677278 como un elemento regulador. Los ensayos de luciferasa sugieren que BV677278 es capaz de modular la expresión del promotor de *DCDC2*, pero podría regular otros genes (15). Un haplotipo de riesgo de tres marcadores que abarca la mitad de 5' y la secuencia corriente arriba de *KIAA0319* se ha asociado de manera consistente a un rendimiento reducido en la lectura (21-24). Además, la expresión de *KIAA0319* en líneas celulares neuronales humanas se reduce con este haplotipo, en relación con los haplotipos sin riesgo (25). Por lo tanto, los solicitantes cuestionaron si BV677278 podría interactuar genéticamente con el haplotipo de riesgo de *KIAA0319*, y examinaron el efecto de portar tanto un haplotipo de riesgo de *DCDC2* (CGCGAG o GACGAG) como el de *KIAA0319* en varias mediciones de la lectura, del lenguaje y cognitivas. Sorprendentemente, los sujetos con haplotipos de riesgo en ambos genes mostraron un rendimiento promedio notablemente peor (hasta 0,40 desviaciones estándar) en casi todas las mediciones examinadas (Fig. 3A). Esta reducción en el rendimiento en portadores de ambos haplotipos de riesgo es, para la mayoría de los fenotipos examinados, mayor que la suma de los portadores individuales, lo que indica una interacción sinérgica entre estos dos genes. Este resultado corrobora un informe anterior, que proporcionó evidencia estadística de que *DCDC2* y *KIAA0319* interactúan para influir en el riesgo de DL (26).

Discusión

Dada la notable similitud del exoma humano con los de otros primates superiores, se ha planteado la hipótesis de que los elementos reguladores en rápida evolución son responsables de las grandes diferencias fenotípicas que observamos. Los resultados recientemente publicados del ENCODE Consortium, que mostraron que la mayor parte del genoma no codificante está activa y la mayor parte de la proporción activa es reguladora, prestan un respaldo circunstancial a esta hipótesis (27). Aquí, los solicitantes proporcionan evidencia de tal elemento regulador que afecta la lectura y el lenguaje, dos fenotipos exclusivamente humanos. BV677278 se expandió rápidamente del gorila al ser humano, aunque la secuencia que lo flanquea está bastante conservada (Fig. 1B), y su presencia, longitud y secuencia varían ampliamente entre las especies de primates (Fig. S2).

Este elemento se une específicamente a ETV6, un factor de transcripción, y un protooncogén también conocido como TEL (translocación ETS leucemia). El gen *ETV6* forma fusiones oncogénicas, a menudo con el protooncogén *AML1*, que se ven con frecuencia en la leucemia (28). El efecto de ETV6 en la transcripción es generalmente represivo, a través del reclutamiento de un complejo correpresor (29). El ETV6 monomérico esencialmente no tiene afinidad por su secuencia de unión; debe al menos dimerizarse para unirse al ADN (30). Existe evidencia de que ETV6 se polimeriza *in vivo*, y la longitud del polímero depende del número y el espaciado de los sitios de unión (31). Esta propiedad sugiere que diferentes alelos de BV677278 se unen a polímeros de ETV6 de diferentes longitudes, dependiendo del número de sitios de unión de ETV6 con un espaciado adecuado, y que estas diferencias cambian la potencia reguladora del complejo (Fig. 3B). Respaldada esta idea la similitud estructural de los alelos 5 y 6: ambos tienen la misma inserción de GGAA en la unidad de repetición 2, en relación con el alelo más común (Tabla S4). GGAA es la secuencia de unión principal de ETV6 (30), y esta inserción podría reclutar un monómero de ETV6 adicional al complejo.

Sin embargo, es incierto si ETV6 reprime la transcripción en este contexto y a qué genes se dirige. Los solicitantes informaron previamente de que los ensayos de luciferasa parecen indicar que algunos alelos de BV677278 activan la transcripción del promotor de *DCDC2*, y que los alelos con estructuras muy diferentes (por ejemplo, 3 y 5, Tabla S4) activan la transcripción en un grado similar (15). La interacción genética de BV677278 con el haplotipo de riesgo de *KIAA0319*, y su dramático efecto sobre el fenotipo, sugieren que *KIAA0319* es un gen objetivo *in vivo*. Se sabe que el haplotipo de riesgo de *KIAA0319* está asociado a una expresión reducida de *KIAA0319*, al menos en líneas celulares neuronales humanas, lo que sugiere la posibilidad de que porte un promotor o variante proximal del promotor que aumente la represión (o disminuya la activación) por BV677278, lo que da como resultado una expresión génica reducida y posibles consecuencias fenotípicas. Esa reducción en el coeficiente intelectual también se observó con la interacción *DCDC2-KIAA0319* (Fig. 3A), puede reflejar una patología a nivel celular (por ejemplo, migración neuronal interrumpida), o simplemente puede reflejar la importancia del lenguaje en la medición del coeficiente intelectual. El genotipado de BV677278 en todos los miembros de ALSPAC y el posterior análisis combinatorio, junto con los experimentos de conformación de la cromatina, arrojarán aún más luz sobre los genes objetivo y el mecanismo de acción de BV677278.

Los haplotipos de riesgo de *DCDC2* y *KIAA0319* tienen un efecto sinérgico en fenotipos de la lectura, el lenguaje y la cognición. Esto confiere credibilidad a la hipótesis de "heredabilidad fantasma", que explica la llamada heredabilidad perdida de rasgos continuos como resultado de interacciones no aditivas entre variantes de riesgo (32). También respalda esta idea el hecho de que aunque los portadores de los haplotipos de riesgo de *DCDC2* muestran un rendimiento promedio reducido en las medidas fenotípicas, las desviaciones estándar para estas medidas fueron generalmente similares a las de los no portadores (Tabla 2). Esto implica que la magnitud del efecto de los haplotipos de riesgo sobre el fenotipo yace en un continuo y depende de otras variantes de riesgo que interactúan, así como de factores ambientales y estocásticos. Además, estos resultados pueden explicar parcialmente la eficacia faltante de los estudios GWAS. Si los elementos reguladores que evolucionan rápidamente son, de hecho, sustancialmente responsables de la variación fenotípica continua, se esperaría que, como BV677278, muestren un mayor grado de polimorfismo que el SNP promedio. Esto los haría difíciles de identificar mediante un análisis estándar de marcadores únicos en GWAS, reforzando la importancia de los análisis de interacción de genes, marcadores múltiples y vías en el estudio de rasgos complejos.

Referencias

1. Development NLoCHA (2010) Learning Disabilities (nichd.nih.gov/health/topics/learning_disabilities.cfm).
2. Peterson RL y Pennington BF (2012) *Lancet* 379(9830):1997-2007.
3. Pennington BF y Bishop DV (2009) *Annual review of psychology* 60:283-306.
4. Anónimo (2007) *National Assessment of Educational Progress (NAEP): The Nation's Report Card, Reading 2007* (National Center for Education Statistics, (nces.ed.gov/nationsreportcard/pubs/main2007/2007496.asp), (Institute of Education Sciences UDoE).
5. Scerri TS y Schulte-Korne G (2010) *European child & adolescent psychiatry* 19(3):179-197.
6. Meng H, et al. (2005) *Human genetics* 118(1):87-90.
7. Peschansky VJ, et al. (2010) *Cereb Cortex* 20(4):884-897.

8. Poelmans G, *et al.* (2011) *Molecular psychiatry* 16(4):365-382.
9. Liu JS (2011) *Current neurology and neuroscience reports* 11(2):171-178.
10. Velayos-Baeza A, *et al.* (2010) *The Journal of biological chemistry* 285(51):40148-40162.
- 5 11. Meng H, *et al.* (2005) *Proceedings of the National Academy of Sciences of the United States of America* 102(47):17053-17058.
12. Ludwig KU, *et al.* (2008) *Psychiatric genetics* 18(6):310-312.
13. Marino C, *et al.* (2012) *Psychiatric genetics* 22(1):25-30.
14. Wilcke A, *et al.* (2009) *Annals of dyslexia* 59(1):1-11.
15. Meng H, *et al.* (2011) *Behavior genetics* 41(1):58-66.
- 10 16. Cope N, *et al.* (2012) *NeuroImage* 63(1):148-156.
17. Boyd A, *et al.* (2012) *International journal of epidemiology*.
18. Bishop DV, *et al.* (1996) *Journal of child psychology and psychiatry, and allied disciplines* 37(4):391-403.
19. Newbury DF, *et al.* (2009) *American journal of human genetics* 85(2):264-272.
20. Mittler G, *et al.* (2009) *Genome research* 19(2):284-293.
- 15 21. Francks C, *et al.* (2004) *American journal of human genetics* 75(6):1046-1058.
22. Luciano M, *et al.* (2007) *Biological psychiatry* 62(7):811-817.
23. Paracchini S, *et al.* (2008) *The American journal of psychiatry* 165(12):1576-1584.
24. Scerri TS, *et al.* (2011) *Biological psychiatry* 70(3):237-245.
25. Paracchini S, *et al.* (2006) *Human molecular genetics* 15(10):1659-1666.
- 20 26. Harold D, *et al.* (2006) *Molecular psychiatry* 11(12):1085-1091, 1061.
27. Djebali S, *et al.* (2012) *Nature* 489(7414):101-108.
28. Fuka G, *et al.* (2011) *PloS one* 6(10):e26348.
29. Wang L y Hiebert SW (2001) *Oncogene* 20(28):3716-3725.
30. Green SM, *et al.* (2010) *The Journal of biological chemistry* 285(24):18496-18504.
- 25 31. Kim CA, *et al.* (2001) *The EMBO journal* 20(15):4173-4182.
32. Zuk O, *et al.* (2012) *Proceedings of the National Academy of Sciences of the United States of America* 109(4):1193-1198.
33. Han B, *et al.* (2008) *Annals of human genetics* 72(Pt 6):834-847.
34. Barrett JC, *et al.* (2005) *Bioinformatics* 21(2):263-265.
- 30 35. Purcell S, *et al.* (2007) *American journal of human genetics* 81(3):559-575.
36. Butter F, *et al.* (2010) *EMBO reports* 11(4):305-311.
- Referencias suplementarias
- S1. Han B, *et al.* (2008) *Annals of human genetics* 72(Pt 6):834-847.
- S2. Scerri TS, *et al.* (2011) *Biological psychiatry* 70(3):237-245.
- 35 S3. Barrett JC, *et al.* (2005) *Bioinformatics* 21(2):263-265.
- S4. Purcell S, *et al.* (2007) *American journal of human genetics* 81(3):559-575.
- S5. Wu KK (2006) *Methods Mol Biol* 338:281-290.

S6. Dignam JD, *et al.* (1983) *Nucleic acids research* 11(5):1475-1489.

S7. Mittler G, *et al.* (2009) *Genome research* 19(2):284-293.

S8. Meng H, *et al.* (2011) *Behavior genetics* 41(1):58-66.

S9. Kolodziej KE, *et al.* (2009) *BMC molecular biology* 10:6.

5 S10. Meng H, *et al.* (2005) *Proceedings of the National Academy of Sciences of the United States of America* 102(47):17053-17058.

S11. Boyd A, *et al.* (2012) *International journal of epidemiology*.

S12. Hulme C, *et al.* (2007) *Paired-associate learning, phoneme awareness, and learning to read. Journal of experimental child psychology* 96(2): 150-166.

10 S13. Peterson RL y Pennington BF (2012) *Lancet* 379(9830):1997-2007.

S14. Rosner J y Simon DP (1971) *Journal of Learning Disabilities* 4(384):40-48.

S15. Rust J, *et al.* (1993) *WORD: Wechsler Objective Reading Dimensions Manual* (Psychological Corporation, Sidcup, Reino Unido).

S16. Nunes T, *et al.* (2003) *Scientific Studies of Reading* 7(3):289-307.

15 S17. Neale MD (1997) *Neale Analysis of Reading Ability, Revised: Manual for Schools* (NFER-Nelson).

S18. Gathercole SE y Baddeley AD (1996) *The Children's Test of Nonword Repetition* (Psychological Corporation, Londres, Reino Unido).

S19. Wechsler D (1996) *Wechsler Objective Language Dimensions (WOLD)* (Psychological Corporation, Londres, Reino Unido).

20 S20. Bishop DV, *et al.* (1996) *Journal of child psychology and psychiatry, and allied disciplines* 37(4):391-403.

S21. Newbury DF, *et al.* (2009) *American journal of human genetics* 85(2):264-272.

Tablas del Ejemplo 1

25 Tabla 1: Datos de asociación y desequilibrio de enlace para haplotipos de riesgo de *DCDC2*. Los fenotipos se describen en la Tabla S1 y en "Materiales y métodos". Los casos se definen con una puntuación menor o igual a 2SD por debajo de la media. Los valores P que sobrevivieron a la corrección de Bonferroni para pruebas múltiples ($\alpha = 0,05$) están en negrita. "% Alelo 5", etc. significa "porcentaje de portadores de haplotipos con al menos una copia de ese alelo o grupo de alelos". Clado 1, la rama filogenética de alelos que incluye 5 y 6, se describe en la Fig. S1.

Datos de asociación						
Haplotipo	Fenotipo -2SD	n casos	n controles	Haplotipo freq.	Ratio de probabilidades	Valor p
CGCGAG	Elim. de fonemas (DL)	89	5225	0,0236	3,20	6,068 x 10⁻⁵
GACGAG	WOLD/NWR (TL)	270	5240	0,0364	1,91	284 x 10 ⁻⁴
Datos de DA						
Haplotipo	n portadores	% Alelo 5	% Alelo 6	% Clado 1		
CGCGAG	226	92,0	7,5	94,3		
GACGAG	392	12,0	77,6	91,3		

30 Tabla 2: Rendimiento medio en mediciones de la lectura y cognitivas en portadores de haplotipos de riesgo de *DCDC2* en comparación con no portadores. La desviación estándar se muestra entre paréntesis al lado de cada

ES 2 718 343 T3

media. El número de sujetos en cada categoría se muestra debajo de esa categoría. Los valores P son de las pruebas T de Student que comparan los medios de portadores y no portadores de cada haplotipo; los valores de p inferiores a 0,05 están marcados con un asterisco. NWR/WOLD se refiere a la puntuación z promedio del rendimiento en las tareas de comprensión verbal de NWR y WOLD. Los fenotipos se describen en la Tabla S1 y en "Materiales y métodos".

5

	Haplotipo CGCGAG (DL)			Haplotipo GACGAG (TL)		
	Portadores	No portadores	Valor p	Portadores	No portadores	Valor p
Lectura 7	27,34 (9,04)	29,01 (8,77)	0,005*	29,09 (8,62)	28,92 (8,80)	0,728
N	232	4929		358	4803	
Ortografía 7	24,38 (13,46)	26,29 (12,33)	0,023*	25,56 (12,77)	26,26 (12,36)	0,305
N	229	4896		355	4770	
Fonema	19,30 (10,00)	20,80 (9,17)	0,016*	20,61 (9,20)	20,74 (9,21)	0,796
N	230	4909		357	4782	
Lectura 9	7,37 (2,71)	7,73 (2,27)	0,020*	7,75 (2,33)	7,72 (2,29)	0,754
N	228	4914		359	4783	
NW Lectura 9	5,05 (2,58)	5,38 (2,36)	0,043 *	5,47 (2,36)	5,36 (2,44)	0,391
N	228	4911		359	4780	
Ortografía 9	10,03 (2,58)	10,50 (3,23)	0,031*	10,48 (3,25)	10,48 (3,26)	0,987
N	228	4904		357	4775	
Velocidad	105,44 (11,76)	106,34 (12,10)	0,299	106,71 (11,77)	106,27 (12,11)	0,524
N	207	4430		326	4311	
Precisión	102,77 (14,00)	105,22 (13,10)	0,009*	105,18 (13,24)	105,11 (13,15)	0,919
N	208	4438		329	4317	
Comp. de lectura	99,74 (11,67)	101,54 (11,37)	0,026*	101,73 (11,82)	101,44 (11,35)	0,663
N	208	4438		329	4317	
CI verbal	107,35 (15,70)	108,97 (15,67)	0,113	108,38 (15,90)	108,94 (15,65)	0,497
N	245	5334		388	5191	
CI de rend.	101,23 (14,96)	100,28 (16,16)	0,366	101,10 (15,72)	101,19 (16,14)	0,913
N	245	5334		388	5191	
CI total	104,58 (14,22)	106,05 (15,26)	0,138	105,62 (14,95)	106,01 (15,23)	0,623

ES 2 718 343 T3

	Haplotipo CGCGAG (DL)			Haplotipo GACGAG (TL)		
	Portadores	No portadores	Valor p	Portadores	No portadores	Valor p
N	245	5334		388	5191	
NWR	7,54 (1,94)	7,58 (1,91)	0,724	7,40 (1,91)	7,55 (1,91)	0,136
N	245	5276		384	5137	
WOLD	7,11 (2,56)	7,33 (2,44)	0,178	7,12 (2,60)	7,33 (2,43)	0,104
N	245	5270		383	5132	
NWR/WOLD	-0,031 (0,82)	0,00 (0,78)	0,532	-0,08 (0,77)	0,01 (0,78)	0,041*
N	245	5281		384	5142	

Tabla S1: (A) Lista de fenotipos utilizados. En los "Materiales y métodos" se proporciona una descripción detallada de cada fenotipo. (B) Definiciones de casos/controles utilizados en el análisis de asociación.

A.	
Fenotipo	Descripción
Lectura a los 7	Dimensiones Objetivas de la Lectura de Wechsler (WORD), tarea de lectura de palabras sueltas, 7 años de edad
Lectura a los 9	Tarea de lectura de palabras sueltas, 9 años de edad
Elim. de fonemas	Tarea de análisis auditivo, 7 años de edad
CI total	Escala de Inteligencia de Wechsler para Niños (WISC), CI total, 8 años de edad
CI verbal	Componente de CI Verbal de la WISC, 8 años de edad
CI de rendimiento	Componente de CI de Rendimiento de la WISC, 8 años de edad
WOLD	Dimensiones Objetivas del Lenguaje de Wechsler (WOLD), tarea de comprensión verbal, 8 años de edad
NWR	Tarea de repetición de seudopalabras, 8 años de edad
Lectura NW a los 9	Tarea de lectura de seudopalabras, 9 años de edad
Ortografía a los 7	Tarea de ortografía con palabras sueltas, 7 años de edad
Ortografía a los 9	Tarea de ortografía con palabras sueltas, 9 años de edad

A.	
<u>Fenotipo</u>	<u>Descripción</u>
Velocidad	Lectura de textos, velocidad, 9 años de edad
Precisión	Lectura de textos, precisión de las palabras leídas, 9 años de edad
Comp. lectora	Lectura de textos, comprensión, 9 años de edad
B.	
<u>Fenotipo</u>	<u>Descripción</u>
Elim. de fonemas	Los casos definidos tienen una puntuación menor o igual a 2 desviaciones estándar por debajo de la media en la tarea de eliminación de fonemas.
WOLD/NWR	Los casos definidos tienen una puntuación menor o igual a 2 desviaciones estándar por debajo de la media en la tarea de comprensión verbal de WOLD o en la tarea de repetición de pseudopalabras.

5

Tabla S2: Correlaciones bivariadas de Pearson entre las mediciones de la lectura y el lenguaje en ALSPAC. Fonema = Tarea de eliminación de fonemas a los 7 años de edad; NWR = Repetición de pseudopalabras a los 8 años de edad; WOLD = Tarea de comprensión verbal de las Dimensiones Objetivas del Aprendizaje de Wechsler a los 8 años de edad; Prom. NWR WOLD = promedio del rendimiento de la puntuación z en las tareas de NWR y WOLD; Lectura 7 = Lectura de palabras sueltas a los 7 años de edad; Lectura 9 = Lectura de palabras sueltas a los 9 años de edad; Todas las mediciones del CI fueron recopiladas a los 8 años de edad con la Escala de Inteligencia para Niños de Wechsler, versión III.

	Fonema	NWR	Comp. WOLD	de	Prom. WOLD	NWR	Lectura 7	Lectura 9	CI total	CI verbal	CI de rend.
Fonema	1										
NWR	0,362	1									
WOLD	0,165	0,214	1								
Prom. NWR	0,338	0,779	0,780		1						
WOLD											
Lectura 7	0,688	0,403	0,259		0,425		1				
Lectura 9	0,550	0,351	0,202		0,355		0,722	1			
CI total	0,406	0,324	0,386		0,455		0,500	0,387	1		
CI verbal	0,426	0,346	0,424		0,494		0,536	0,421	0,871	1	
CI de rend.	0,246	0,192	0,216		0,262		0,292	0,218	0,819	0,435	1

10 Tabla S3: Estadísticas para los SNP reportados en el presente. Las estadísticas se calcularon después de la exclusión de muestras con una tasa de llamada baja (<85 % de la tasa de llamada promedio) y de individuos que no

ES 2 718 343 T3

eran de ascendencia europea. Los SNP en fuente normal (no en negrita) (SNP 1-6 en la tabla) comprenden el bloque de haplotipos de riesgo de *DCDC2*; los SNP en letra negrita (SNP 7-9 en la tabla) comprenden el bloque de haplotipos de riesgo de *KIAA0319*. Los SNP se enumeran en el orden de sus respectivos haplotipos (por ejemplo, CGCGAG).

SNP	Tasa de llamada	Alelo mayor	Frec. de alelo mayor	Alelo menor	Frec. de alelo menor	Valor p HWE
rs33914824	92,6 %	C	0,961	G	0,039	0,541
rs807694	94,0 %	G	0,952	A	0,047	0,974
rs707864	93,0 %	T	0,874	C	0,126	0,012
rs10456301	93,5 %	G	0,929	A	0,071	0,814
rs16889066	91,2 %	A	0,945	G	0,055	0,134
rs9379651	86,7 %	G	0,877	A	0,123	0,720
rs4504469	89,1 %	C	0,592	T	0,408	0,054
rs2038137	90,0 %	G	0,630	T	0,370	0,611
rs2143340	89,6 %	A	0,849	G	0,151	0,583

5

Tabla S4: Estructuras y frecuencias poblacionales para todos los alelos de BV677278 descritos hasta la fecha. Las frecuencias alélicas para los alelos disponibles se calcularon a partir de un estudio previo (10). Se encuentran disponibles frecuencias alélicas poblacionales solo para los alelos 11-22 en portadores de haplotipos de riesgo de *DCDC2* (véase la Tabla S5). “Del” se refiere a la microdelección de 2.445 pb que abarca a BV677278.

Alelo	Unidad de repetición 1 SEQ ID NO: 15 y 16	Unidad de repetición 2	Unidad de repetición 3	Unidad de repetición 4	Región constante	Unidad de repetición 5	Región constante	Frec. de alelo*	Frec. de alelo**	Longitud
1	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,624	0,5536	102
2	(GAGAGGAAGGAAA)1 (SEQ ID NO:16)	(GGAA)9 (SEQ ID NO:20)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,003	0,0143	85
3	(GAGAGGAAGGAAA)1 (SEQ ID NO:16)	(GGAA)6 (SEQ ID NO:21)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,060	0,0464	85
4	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)6 (SEQ ID NO:21)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,106	0,1429	98
5	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)8 (SEQ ID NO:22)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,028	0,0143	106
6	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)8 (SEQ ID NO:22)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	0,039	0,0571	102
7	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)8 (SEQ ID NO:22)	(GAAA)1	(GGAA)1	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,003	0	102

ES 2 718 343 T3

Alelo	Unidad de repetición 1 SEQ ID NO: 15 y 16	Unidad de repetición 2	Unidad de repetición 3	Unidad de repetición 4	Región constante	Unidad de repetición 5	Región constante	Frec. de alelo*	Frec. de alelo**	Longitud
		NO:22)				NO:23)				
8	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,003	0	90
9	(GAGAGGAAGGAAA) 1 (SEQ ID NO:16)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,005	0,0179	89
10	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)4 (SEQ ID NO:23)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	0,044	0,0286	90
11	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	0	98
12	(GAGAGGAAGGAAA)1 (SEQ ID NO:16)	(GGAA)8 (SEQ ID NO:22)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	0,0036	89
13	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)9 (SEQ ID NO:20)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	0,0071	106
14	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)9 (SEQ ID NO:20)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	110
15	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)5 (SEQ ID NO:24)	(GAAA)2	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	98
16	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)5 (SEQ ID NO:24)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	94 *Solo placa Coriell AfA
17	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)4 (SEQ ID NO:23)	(GAAA)2	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	94 *Solo placa Coriell AfA
18	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)2	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	106 *Solo placa Coriell AfA
19	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)9 (SEQ ID NO:20)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	98
20	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)10 (SEQ ID NO:25)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	114
21	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)6 (SEQ ID NO:21)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	N/A	94
22	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)10 (SEQ ID NO:25)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102

ES 2 718 343 T3

Alelo	Unidad de repetición 1 SEQ ID NO: 15 y 16	Unidad de repetición 2	Unidad de repetición 3	Unidad de repetición 4	Región constante	Unidad de repetición 5	Región constante	Frec. de alelo*	Frec. de alelo**	Longitud
23	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)11 (SEQ ID NO:26)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	106
24	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)6 (SEQ ID NO:21)	(GAAA)2	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102
25	(GAGAGGAAGGAAA)1 (SEQ ID NO:16)	(GGAA)8 (SEQ ID NO:22)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	93
26	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)5 (SEQ ID NO:24)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	N/A	90
27	(GAGAGGAAGGAAA)1 (SEQ ID NO:16)	(GGAA)5 (SEQ ID NO:24)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	81
28	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)5 (SEQ ID NO:24)	(GGGA)2	N/A	N/A	106
29	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)5 +(GGGA)1 +(GGAA)1 (SEQ ID NO:27)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102
30	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)5 (SEQ ID NO:24)	(GAAA)1	(GGAA)4 (SEQ ID NO:23)	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102
31	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)1 +(GGGA)1	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102
32	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)8 (SEQ ID NO:22)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	94
33	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)6 (SEQ ID NO:21)	(GAAA)0	(GGAA)0	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	N/A	82
34	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)7 (SEQ ID NO:19)	(GAAA)2	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)3 (SEQ ID NO:29)	(GGGA)2	N/A	N/A	102
35	(GAGAGGAAGGAAA) -(GAGAGGAAGGAAA) (SEQ ID NO:17)	(GGAA)7 (SEQ ID NO:19)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	102
36	(GAGAGGAAGGAAA) -(GAGAGGAAGGAAA) (SEQ ID NO:18)	(GGAA)9 (SEQ ID NO:20)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)4 (SEQ ID NO:23)	(GGGA)2	N/A	N/A	109
37	(GAGAGGAAGGAAA)2 (SEQ ID NO:15)	(GGAA)6 (SEQ ID NO:21)	(GAAA)1	(GGAA)2	GGAAAGAATGAA (SEQ ID NO:28)	(GGAA)1 +(GAAA)1 +(GGAA)2	(GGGA)2	N/A	N/A	98

ES 2 718 343 T3

Alelo 31 (SEQ ID NO: 61)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
GAGGAAAGAATGAAGGAAGGAAGGAAGGAAGGGAGGGA

Alelo 32 (SEQ ID NO: 62)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AATGAAGGAAGGAAGGAAGGAAGGGAGGGA

5 Alelo 33 (SEQ ID NO: 63)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AAGGAAGGAAGGGAGGGA

Alelo 34 (SEQ ID NO: 64)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AAGGAAGGAAGGAATGAAGGAAGGAAGGAAGGGAGGGA

Alelo 35 (SEQ ID NO: 65)

10 GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AAGGAAAGAATGAAGGAAGGAAGGAAGGAAGGGAGGGA

Alelo 36 (SEQ ID NO: 66)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AGGAAGGAAGGAAGGAATGAAGGAAGGAAGGAAGGAAGGGAGGGA

Alelo 37 (SEQ ID NO: 67)

GAGAGGAAGGAAAGAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AAAGAATGAAGGAAGGAAGGAAGGAAGGGAGGGA

15 Alelo 38 (SEQ ID NO: 68)

GAGAGGAAGGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG
AGGAAGGAAGGAAGGAAGGGAGGGA

Tabla S5: Frecuencias de alelos de BV677278 en los portadores de los haplotipos CGCGAG y GACGAG en ALSPAC. Los alelos que pertenecen al Clado 1 están demarcados con un "1" entre paréntesis.

Alelo	Frecuencia de alelos de BV677278, CGCGAG	Frecuencia de alelos de BV677278, GACGAG
1	0,323	0,367
2	0	0
3	0,024	0,029
4	0,055	0,048
5 (1)	0,473	0,061
6 (1)	0,038	0,393
7	0	0
8	0	0,001
9	0,002	0,004
10	0,027	0,022
11 (1)	0	0,008
12 (1)	0	0

Alelo	Frecuencia de alelos de BV677278, CGCGAG	Frecuencia de alelos de BV677278, GACGAG
13 (1)	0	0,009
14 (1)	0,007	0,009
15	0,002	0
16	0	0
17	0	0
18	0	0,001
19	0	0
20 (1)	0	0,005
21 (1)	0	0,003
22	0	0,001
Del	0,049	0,037

5 Tabla S6: Valores p para el experimento ChIP-qPCR (Figura 2B). Los valores representan pruebas T emparejadas de una cola para el enriquecimiento multiplicado entre cada par de condiciones de ChIP especificadas (tres réplicas de cada una). STR: Amplicón STR de BV677278; β -actina: Amplicón de control del gen de la β -actina; α -ETV6: ChIP con anticuerpo anti-ETV6; α -H3: ChIP con anticuerpo de control antivariante de histona H3; NA: ChIP sin control de anticuerpos. Los valores p por debajo de 0,05 se muestran en negrita con un asterisco, los valores p por debajo de 0,01 tienen dos asteriscos.

	α -ETV6, STR	α -ETV6, β -actina	NA, STR	α -H3, STR	α -H3, β -actina	NA, β -actina
α -ETV6, STR	X	0,00728**	0,01114*	0,07659	0,06967	0,01005*
α -ETV6, β -actina		X	0,19296	0,05375	0,05228	0,11426
NA, STR			X	0,02079*	0,08653	0,42981
α -H3, STR				X	0,46893	0,01133*
α -H3, β -actina					X	0,05509
NA, β -actina						X

10 Tabla S7: ANOVA de una vía entre los grupos enumerados en la Figura 3A: portadores de 1) haplotipo sin riesgo, 2) haplotipo de riesgo de *KIAA0319*, 3) haplotipo de *DCDC2*, y 4) un haplotipo de riesgo en ambos genes.

Fenotipo (transformado en Z)	Valor p de ANOVA
Eliminación de fonemas	0,088

Fenotipo (transformado en Z)	Valor p de ANOVA
CI total	0,003
CI verbal	0,006
CI de rendimiento	0,043
Comprensión verbal (WOLD)	0,562
Repetición de seudopalabras	0,053
Prom. WOLD y NWR	0,179

Ejemplo 2

Una vez establecida la asociación entre los haplotipos de riesgo de *DCDC2* y los haplotipos de riesgo de *KIAA0319*, se investigó la relación de READ1 (una variante funcional de riesgo de DL/TL dentro de *DCDC2*) y la variante de riesgo de DL de *KIAA0319*, KIAHap (un haplotipo que cubre la mitad 5' del gen de riesgo de DL *KIAA0319*, así como parte de su secuencia corriente arriba y el gen vecino *TDP2*). Como se describe en el presente documento, READ1 está compuesto por cinco unidades de repetición discretas, cada una de las cuales varía en número, lo que da lugar a un polimorfismo considerable. Incluyendo la delección, se han descrito 39 alelos de READ1 hasta el momento, 6 de los cuales son comunes y 32 de los cuales son raros (en individuos de ascendencia europea). Estos alelos varían en longitud, de 81 pb a 114 pb. READ1 no parece existir fuera de los primates superiores, y entre las especies de primates superiores (y dentro del *Homo sapiens*) su longitud es muy variable. Parece que READ1 es un elemento hipermutable y de rápida evolución que apareció por primera vez en primates y alcanzó su tamaño completo en el *Homo sapiens* (Powers *et al.*, 2013).

Para determinar los efectos de los alelos individuales de READ1 *in vivo*, se examinó la asociación de READ1 a la lectura y el lenguaje en el Estudio Longitudinal Avon de Padres e Hijos (ALSPAC), una gran cohorte de nacimiento longitudinal. El alelo 5 de READ1 se asoció fuertemente a DL grave, mientras que el alelo 6 de READ1 se asoció fuertemente a TL grave (Tabla 1). Además, aquellos individuos que portaban al menos una copia del alelo 5 tuvieron un peor desempeño en seis mediciones diferentes relacionadas con la lectura, en promedio, en comparación con los no portadores del alelo 5 ($p < 0,05$), mientras que los portadores del alelo 6 tuvieron un peor desempeño en una medición compuesta del lenguaje, en promedio, que los no portadores del alelo 6 ($p < 0,05$). Estos dos alelos relativamente comunes (frecuencias alélicas del 3,6 % y 5,0 %, respectivamente) son estructuralmente similares entre sí, y se agrupan filogenéticamente en el mismo clado. En comparación con el alelo más común (alelo 1), ambos alelos tienen una inserción GGAA en la misma posición. Por el contrario, otros alelos de READ1 comunes parecían tener un efecto protector en la lectura y el lenguaje. Los portadores de al menos una copia de un alelo de READ1 más corto (un alelo con una sola iteración de la unidad de repetición 1 o RU1, a diferencia de dos iteraciones en la mayoría de los alelos de READ1) obtuvieron mejores resultados en las tareas de lectura y del lenguaje en comparación con los no portadores, aunque las asociaciones protectoras a DL y TL graves solo fueron sugerentes (Tabla 1 del Ejemplo 2, más adelante).

Tabla 1: Resultados de la asociación para READ1. Esta tabla muestra los resultados de la asociación de los alelos simples y compuestos de READ1 con DL grave y TL grave. Las asociaciones se calcularon según un modelo alélico. Los valores p significativos están en negrita, y se muestran los ratios de probabilidades notables para el riesgo aumentado ([#]) y reducido (*).

Alelo	Descripción	Alelos de READ1	FAM (ALSPAC)	Valor p, DL grave	OR, DL grave	Valor p, TL grave	OR, TL grave
3	Solo 3, protector	3	0,0463	0,179	0,575*	0,255	0,77
4	Solo 4	4	0,0924	0,239	1,28	0,141	0,78
5	Solo 5, deletéreo	5	0,0355	0,000058	2,37[#]	0,487	0,84
6	Solo 6, deletéreo	6	0,0496	0,0995	1,53[#]	0,00595	1,65

Alelo	Descripción	Alelos de READ1	FAM (ALSPAC)	Valor p, DL grave	OR, DL grave	Valor p, TL grave	OR, TL grave
10	Solo 10	7	0,0502	0,795	0,919	0,603	0,9
5+6	Alelos deletéreos principales	5,6	0,0851	0,000037	1,96[#]	0,000074	1,73
Alelos más cortos	Solo una copia de RU1; alelos protectores principales	2, 3, 9, 12, 25, 27, 38	0,0521	0,0957	0,506*	0,292	0,8

Como se describe en el presente documento, la proteína de unión a READ1 se identificó como el potente represor transcripcional ETV6, cuya secuencia de unión (GGAA) coincide con la inserción observada en los alelos de riesgo 5 y 6. Esto sugiere que los alelos 5 y 6 tienen más sitios de unión a ETV6 disponibles que los alelos neutros o protectores (particularmente los alelos protectores más cortos). Cuando se consideran las propiedades de ETV6, el comportamiento peculiar de los alelos de READ1 comienza a tener sentido. En su estado monomérico, ETV6 es incapaz de unirse al ADN. Debe al menos homodimerizarse para desplazar un dominio autoinhibitorio que bloquea su dominio de unión al ADN en el estado monomérico (Green *et al.*, 2010), y se sabe que es capaz de homopolimerizarse (Tognon *et al.*, 2004). Esta propiedad sugiere una posibilidad intrigante: los alelos de READ1 de diferentes longitudes se unen a diferentes cantidades de monómeros ETV6, y esto altera la potencia reguladora del complejo (Figura 3B).

Asimismo, se cree que KIAHap marca un promotor funcional o una variante proximal del promotor que altera la regulación génica de *KIAA0319*. Hay evidencia de que esta variante funcional es el SNP rs9461045. Este SNP está en desequilibrio de enlace con KIAHap y está asociado a una expresión reducida del gen reportero del promotor de *KIAA0319* en líneas celulares tanto neuronales como no neuronales (Dennis *et al.*, 2009; Paracchini *et al.*, 2006). La naturaleza reguladora de KIAHap y el hecho de que READ1 parece ser un elemento regulador de la unión a ETV6 llevó a la pregunta de si los alelos de riesgo 5 y 6 de READ1 interactúan con KIAHap para afectar el fenotipo. Sorprendentemente, los portadores de un alelo de riesgo de READ1 y KIAHap mostraron un desempeño notablemente peor en las tareas de lectura, lenguaje y CI (la Figura 5A muestra el alelo 5). Esta interacción es generalmente sinérgica; es decir, el efecto de tener dos alelos de riesgo es mayor que la suma de los efectos individuales de los alelos (Figura 5A). Además, los alelos protectores más cortos de READ1 (por ejemplo, aquellos que comprenden una sola unidad de repetición 1, "RU1_1") parecen interactuar de manera epistática con KIAHap. En individuos con al menos una copia de un alelo RU1_1 más corto de READ1, el efecto deletéreo de KIAHap es completamente denegado. El rendimiento promedio de los sujetos con un alelo RU1_1 y KIAHap generalmente es ligeramente superior al promedio de la población y se asemeja al de los sujetos con un alelo RU1_1 solo (Figura 5B). Por lo tanto, parece que KIAHap sinergiza con alelos de READ1 deletéreos para exacerbar sus efectos deletéreos individuales en la lectura y el lenguaje, pero está enmascarado epistáticamente por alelos protectores de READ1. Esto sugiere que ETV6 y READ1 forman un complejo regulador que regula epistáticamente a *KIAA0319* y, posiblemente, *DCDC2* y otros genes objetivo.

Ejemplo 3

Caracterización del locus DYX2 en el cromosoma 6p22 con la discapacidad lectora, el trastorno del lenguaje y la cognición en general

Introducción

Aquí se describe la evaluación de la relación del locus DYX2 con DL, TL y la cognición. Se desarrolló un panel de marcadores que cubre densamente el locus DYX2 de 1,4 Mb y se usó para evaluar la asociación a las mediciones de lectura, lenguaje y cognitivas en sujetos del Estudio Longitudinal Avon de Padres e Hijos (ALSPAC). Las asociaciones se replicaron luego en tres cohortes independientes seleccionadas. Confirmando los resultados de los otros ejemplos descritos en el presente documento, hubo asociaciones con genes de riesgo de DL conocidos, *KIAA0319* y *DCDC2* (Figura 6A). Además, se identificaron otros marcadores en o cerca de otros genes de DYX2, incluidos *TDP2*, *ACOT13*, *C6orf62*, *FAM65B* y *CMAHP*. La estructura de DA del locus sugiere que los aciertos de asociación dentro de *TDP2*, *ACOT13* y *C6orf62* están capturando una variante de riesgo reportada previamente en *KIAA0319*. Estos resultados respaldan aún más que *KIAA0319* y *DCDC2* son genes efectores principales en DYX2, e identifican a *FAM65B* y *CMAHP* como nuevos genes de riesgo de DYX2. La asociación de DYX2 con múltiples rasgos neuroconductuales sugiere que las variantes de riesgo tienen consecuencias funcionales que afectan a múltiples procesos neurológicos.

Métodos

Sujetos

La cohorte de descubrimiento en esta investigación fue el Estudio Longitudinal Avon de Padres e Hijos (ALSPAC). ALSPAC es una cohorte de nacimientos basada en la población con sede en Avon, Reino Unido. Los sujetos fueron reclutados antes del nacimiento; se reclutaron un total de 15.458 fetos, de los cuales 14.701 estaban vivos al año de edad. El reclutamiento, los participantes y las metodologías de estudio se describen en detalle en otra parte (bristol.ac.uk/alspac) (Boyd *et al.* 2012; Golding *et al.* 2001). Las muestras de ADN para el análisis genético estaban disponibles para 10.259 sujetos. Las mediciones de lectura, lenguaje y cognición se realizaron a los 7, 8 y 9 años de edad. Los sujetos con un CI ≤ 75 en el CI Total de la Escala de Inteligencia de Wechsler para Niños (WISC-III) fueron excluidos para evitar los efectos de confusión de la discapacidad intelectual (Eicher *et al.* 2013a; Eicher *et al.* 2013b; Powers *et al.* 2013; Wechsler *et al.* 2002). Para evitar la estratificación de la población en los análisis genéticos, también se excluyeron los sujetos de ascendencia no europea. Las muestras con tasas de llamada de genotipo generales $< 0,85$ se excluyeron de los análisis, dejando un tamaño de muestra final de 5579 individuos para los análisis de TL y 5525 individuos para los análisis de DL. La aprobación ética se obtuvo del Comité de Ética y Derecho de ALSPAC, los comités locales de ética en investigaciones del Reino Unido y el Comité de Investigación con Seres Humanos de Yale.

Mediciones de la lectura, el lenguaje y la cognición

Las mediciones de la lectura en el ALSPAC utilizadas en esta investigación incluyen una tarea de eliminación de fonemas a los 7 años de edad, la lectura de palabras sueltas a los 7 y 9 años de edad y la lectura de seudopalabras sueltas a los 9 años de edad (Tabla 1a). La tarea de eliminación de fonemas, también conocida como Prueba de Análisis Auditivo, evalúa el reconocimiento de los fonemas, un déficit central en la DL (Rosner y Simon 1971). Para la tarea de eliminación de fonemas, el niño escucha una palabra hablada en voz alta y luego se le pide que elimine un fonema específico de esa palabra para formar una nueva palabra. La lectura de palabras sueltas se evaluó a los 7 años de edad utilizando la subprueba de lectura de las Dimensiones Objetivas de la Lectura de Wechsler (WORD) (Rust *et al.* 1993). A los 9 años de edad, se volvió a evaluar la lectura de palabras sueltas pidiéndole al niño que leyera diez palabras reales y diez seudopalabras en voz alta, un subconjunto de una lista más grande de palabras y seudopalabras (Nunes *et al.* 2003). Para examinar los casos graves (DL grave), se definieron los casos con una puntuación de 2 o más desviaciones estándar por debajo de la media en la tarea de eliminación de fonemas (Tabla 1b). Los casos con DL moderada también se definieron como una puntuación de al menos 1 desviación estándar por debajo de la media en la lectura de palabras sueltas a los 7 años de edad, la lectura de palabras sueltas a los 9 años de edad y la lectura de seudopalabras sueltas los 9 años de edad (Tabla 1b). Se eligió un umbral de 1 desviación estándar, ya que las mediciones se incluyeron en tres puntos de tiempo para aislar a los individuos con habilidades de decodificación persistentemente deficientes. Se examinaron diferentes niveles de gravedad porque los estudios anteriores en el locus DYX2 habían mostrado diferencias en los patrones de asociación genética según la gravedad del caso, particularmente con *KIAA0319* que se asocia a definiciones de casos de DL más moderadas, y a *DCDC2* con definiciones más graves (Paracchini *et al.* 2008; Powers *et al.* 2013; Scerri *et al.* 2011).

Las mediciones de lenguaje se recopilaron a los 8 años de edad (Tabla 1a). Con una adaptación de la Tarea de Repetición de Seudopalabras (NWR), en la que los sujetos repitieron grabaciones de seudopalabras, se evaluó la memoria fonológica a corto plazo y las capacidades de procesamiento (Gathercole y Baddeley 1996). Los niños también realizaron la tarea de comprensión verbal de las Dimensiones Objetivas del Lenguaje de Wechsler (WOLD) a los 8 años de edad (Wechsler *et al.* 1996), en la que respondieron a preguntas sobre un párrafo leído en voz alta por un examinador que describía una imagen presentada. Estas mediciones se analizaron porque se sabe que los individuos con TL muestran un desempeño deficiente en las tareas de NWR y WOLD (Bishop *et al.* 1996; Newbury *et al.* 2009). Al igual que con DL, también se examinó la asociación del locus DYX2 a definiciones de casos moderados y graves de TL. Para evaluar el riesgo impartido para TL grave, los casos de TL grave se definieron como una puntuación de 2 o más desviaciones estándar por debajo de la media de la muestra en cualquiera de las tareas (TL grave) (Tabla 1b). Para evaluar los déficits más moderados, los casos se definieron como puntuaciones de al menos 1,5 desviaciones estándar por debajo de la media de la muestra en cada tarea (NWR moderada y WOLD moderado) (Tabla 1b). Como se usaron menos mediciones para evaluar los rasgos relacionados con TL, el umbral para las definiciones de casos se incrementó a 1,5 desviaciones estándar para evaluar déficits más moderados. Los CI verbal, de rendimiento y total se evaluaron a los 8 años de edad, utilizando la Escala de Inteligencia de Wechsler para Niños (WISC-III) (Tabla 1a). Las mediciones del CI se examinaron como rasgos cuantitativos (Tabla 1b).

Genotipado y análisis genéticos

Se desarrolló un panel de marcadores de SNP para capturar la variación común en el locus DYX2. Los TagSNP en el locus se seleccionaron utilizando el servidor de diseño de estudios de asociaciones de Han *et al.* (Han *et al.* 2008). El panel final de DYX2 contenía 195 SNP con una potencia promedio estimada del 83 % y del 68 % para capturar variantes comunes y raras conocidas, respectivamente, en el locus DYX2 que abarca aproximadamente 1,4 Mb. Los marcadores fueron genotipados en la plataforma Sequenom (San Diego, CA), siguiendo las pautas del fabricante, en el Centro de Análisis de Genomas de Yale (West Haven, CT).

Los marcadores que se desviaban sustancialmente del equilibrio de Hardy-Weinberg, o que tenían una tasa de llamada global <85 %, no se utilizaron para el análisis basado en haplotipos. En la cohorte de descubrimiento de ALSPAC, se llevaron a cabo los análisis de SNP de marcadores únicos de los estados de casos y controles, y los rasgos cuantitativos, utilizando la SNP & Variation Suite (SVS) v7.6.4 (Bozeman, MT). Se evaluó el desequilibrio de enlace y se construyeron bloques de haplotipos utilizando la opción de la regla de cuatro gametos en HaploView v4.2. Las pruebas de asociación de haplotipos se realizaron con haplotipos que tenían frecuencias mayores o iguales al 1 % utilizando PLINK v1.07 (Barrett *et al.* 2005; Purcell *et al.* 2007). Se informa de que las asociaciones con $p < 0,001$ para la cohorte de descubrimiento de ALSPAC presentan resultados sugerentes. Sin embargo, para corregir las pruebas múltiples, se usó un umbral de Bonferroni de 0,000256 (0,05 dividido entre 195 marcadores) para las pruebas de asociación de descubrimiento en la cohorte ALSPAC.

Después de los análisis de descubrimiento en ALSPAC, las variantes asociadas se evaluaron en tres cohortes reclutadas específicamente para DL o TL: TL Iowa, DL Italiana, DL Colorado (Tabla 2). La cohorte TL Iowa está compuesta por 219 casos de TL y 209 controles no relacionados, emparejados por sexo y edad, recopilados en la Universidad de Iowa. Los sujetos se sometieron a varias mediciones del lenguaje, incluidas la prueba Woodcock Johnson-III (WJ) y la Prueba Gray de Lectura Oral (GORT), que se utilizaron para obtener una puntuación compuesta del lenguaje, que luego se dicotomizó en estado caso-control en -1,14 desviaciones estándar (Eicher *et al.* 2013a; Weismer 2000). La cohorte del Centro de Investigación de Discapacidades del Aprendizaje de Colorado (CLDRC) consta de 1201 individuos en 293 familias nucleares. Se reclutaron familias para el estudio si al menos un niño tenía antecedentes de problemas de lectura. La cohorte italiana consta de 878 individuos en 304 familias nucleares; estas familias fueron reclutadas a través de un probando con DL clínicamente diagnosticada. La aprobación ética para el reclutamiento y las metodologías del estudio se obtuvieron del Comité de Investigación con Seres Humanos de Yale y de las Juntas de Revisión Institucional de la Universidad de Iowa, de la Universidad de Denver y de Italia. Los SNP que tenían un solo marcador o asociaciones dentro de haplotipo con $p < 0,001$ en el ALSPAC se probaron para determinar la replicación en las cohortes TL Iowa, DL Italiana y DL Colorado. TL Iowa se analizó utilizando SNP & Variation Suite (SVS) v7.6.4 (Bozeman, MT), mientras que las cohortes DL Italiana y DL Colorado basadas en familias se examinaron utilizando PLINK v1.07 (Purcell *et al.* 2007). Los resultados sugerentes de ALSPAC se derivaron para análisis de replicación a fin de enfatizar la replicación de asociaciones sobre correcciones estadísticas para pruebas múltiples. Se reportan replicaciones con $p < 0,05$ en las cohortes TL Iowa, DL Italiana y DL Colorado.

Resultados

La asociación con marcadores de DYX2 se realizó en tres dominios separados: (1) DL, (2) TL y (3) CI. En aras de la claridad, las asociaciones se presentan dominio por dominio, con un énfasis en la replicación de la fuerza de las asociaciones en lugar de la corrección de pruebas múltiples.

DL

Las asociaciones con DL se realizaron utilizando dos definiciones de gravedad diferentes: (1) DL grave y (2) DL moderada (Tabla 1b). Para DL grave, se encontraron asociaciones con *DCDC2*, *KIAA0319* y *TDP2* (Tabla 3). La asociación de *DCDC2* y DL grave se explora completamente en el Ejemplo 1; Powers *et al.* (2013). El marcador rs2294691 de *TDP2* no replicó su asociación en ninguna de las tres cohortes de replicación (Tabla 5). El marcador rs10456309 de *KIAA0319* se replicó en las cohortes TL Iowa y DL Colorado (Tabla 5). Con DL moderada, hubo una asociación entre rs1562422 cerca del gen *FAM65B* y el pseudogén *CMAHP*, que se replicó en la cohorte DL Colorado (Tabla 3, Tabla 5).

TL

Las pruebas de asociación se realizaron en tres fenotipos de TL: (1) TL grave, (2) NWR moderada y (3) WOLD moderado (Tabla 1b). Al igual que con DL grave, hubo asociaciones entre *DCDC2* y TL grave. El haplotipo de *DCDC2* que se asoció a TL grave se describe en el Ejemplo 1; Powers *et al.* (2013). Un marcador dentro de este haplotipo de *DCDC2*, rs807694, mostró asociación con TL grave y se replicó en la cohorte TL Iowa (Tabla 3, Tabla 5). Con una definición de caso más moderada, se observaron asociaciones a *ACOT13* y *C6orf62* (Tabla 3), genes adyacentes a *KIAA0319* y *TDP2*. Tanto rs3777663 en *ACOT13* como rs3756814 en *C6orf62* mostraron asociaciones en las cohortes DL Italiana y TL Iowa (Tabla 5).

CI

También se realizaron pruebas de asociación entre los marcadores de DYX2 y el CI verbal, el CI de rendimiento y el CI total (Tabla 1b). Las asociaciones del CI verbal incluyeron marcadores únicos y haplotipos que cubren la mitad 5' de *KIAA0319*, rs9348646 en *FAM65B*, y un haplotipo que abarca *ACOT13* y *C6orf62*, con evidencia de replicación (Tabla 4a, Tabla 4b, Tabla 5). Hubo una superposición sustancial de las asociaciones de DYX2 con el CI verbal y las asociaciones con DL y TL. Estas similitudes de asociaciones no son inesperadas, ya que los rasgos están altamente correlacionados y se sabe que capturan dominios similares (Tabla 6). Las asociaciones de DYX2 con el CI de rendimiento y total fueron más débiles; no hubo asociaciones con el CI de rendimiento y una asociación única no

replicada del CI total con rs2328791, que se encuentra en una gran región intergénica telomérica a *NRSN1* y *DCDC2* (Tabla 4a, Tabla 4b, Tabla 5).

Desequilibrio de enlace dentro de *DYX2*

5 En los análisis, se observaron asociaciones replicadas en los siguientes genes: *DCDC2*, *KIAA0319*, *TDP2*, *ACOT13*,
 10 *C6orf62*, *FAM65B* y el pseudogén *CMAHP*. Sin embargo, como estos SNP están muy próximos entre sí, se evaluó el
 15 desequilibrio de enlace (DE) entre el panel de marcadores para determinar si los SNP asociados estaban marcando
 la misma variación en el locus. Como se describe en los Ejemplos 1 y 2, las asociaciones de *DCDC2* marcaban
 alelos de *READ1*. Dentro de *KIAA0319*, parece que hay dos bloques de DE claros que separan el gen en una mitad
 5' y una mitad 3' (datos no mostrados). La mitad 5' de *KIAA0319* está en DE fuerte con *TDP2*, *ACOT13* y *C6orf62*, lo
 que indica que las asociaciones dentro de estos genes pueden capturar la misma variación (datos no mostrados).
 Las asociaciones en *FAM65B* y *CMAHP* parecen estar marcando asociaciones independientes (datos no mostrados).
 Si bien rs1562422 se ubica como intergénico a *FAM65B* y *CMAHP*, este marcador está en DE fuerte con otros
 marcadores dentro del pseudogén *CMAHP*. La integración de los análisis de asociación y la estructura DE indican
 cuatro señales de asociación independientes centradas en (1) *DCDC2*, (2) la mitad 5' de *KIAA0319*, (3) *FAM65B* y (4)
CMAHP.

En este ejemplo, se caracteriza la relación del locus *DYX2* con DL, TL y CI (Figura 6B). Los resultados confirman
 que las asociaciones de los genes de riesgo de DL, *KIAA0319* y *DCDC2*, incluyen TL. Además, *FAM65B* y *CMAHP*
 se identificaron como genes de riesgo para rasgos lingüísticos. Los marcadores dentro del locus *DYX2* mostraron
 asociación con numerosos rasgos de la comunicación, incluidos DL, TL y CI verbal. Hubo una notoria ausencia de
 20 asociaciones de *DYX2* con CI de rendimiento y total, lo que indica que el locus *DYX2* influye en los procesos
 relacionados con el lenguaje en mucho mayor grado que los rasgos cognitivos generales.

La asociación genética de *DYX2* con DL, TL e CI verbal es el último ejemplo de varios procesos neurocognitivos y
 comunicativos que comparten asociaciones genéticas. El solicitante y otros han demostrado que estos rasgos
 neuroconductuales tienen contribuyentes genéticos comunes, que incluyen variantes en *FOXP2*, *KIAA0319*, *CMIP*,
 25 *ZNF385D*, *CNTNAP2* y *DCDC2* (Eicher *et al.* 2013b; Newbury *et al.* 2009; Newbury *et al.* 2011, Pennington y Bishop
 2009, Peter *et al.* 2011; Pinel *et al.* 2012; Powers *et al.* 2013; Scerri *et al.* 2011; Wilcke *et al.* 2012). La expansión de
 la asociación de *DYX2* a partir de la lectura para incluir otros procesos relacionados con el lenguaje sugiere que las
 variantes causales pueden afectar estos rasgos de manera pleiotrópica, en lugar de influir exclusivamente en el
 lenguaje verbal o escrito. Los hallazgos de este estudio coinciden con esta "hipótesis de los genes generalistas", que
 30 también está respaldada por un reciente análisis de rasgos complejos del genoma (GTCA) de las capacidades
 cognitivas y de aprendizaje (Trzaskowski *et al.* 2013). Las fuertes correlaciones y la relación entre estas medidas
 neurocognitivas (Tabla 6) sugieren que estos genes de *DYX2* afectan a procesos centrales del lenguaje que, a su
 vez, se manifiestan fenotípicamente de varias maneras, incluyendo la lectura, el lenguaje y la cognición.

Que los múltiples genes de *DYX2* hayan mostrado una asociación con los fenotipos en este estudio es interesante y,
 a primera vista, algo inesperado. Una posibilidad es que uno o dos genes no sean los únicos responsables de la
 implicación consistente de este locus en los fenotipos de la lectura, el lenguaje y la cognición, como se cree
 principalmente. *KIAA0319* y *DCDC2* se conocen actualmente como los dos genes de mayor riesgo en el locus *DYX2*.
 Ambos han sido implicados tanto en la DL como en la variación subclínica del rendimiento en la lectura, utilizando
 35 mediciones neurocomportamentales clásicas y, más recientemente, técnicas de neuroimagen (Eicher y Gruen 2013;
 40 Graham y Fisher 2013). Otros genes en *DYX2* se han asociado al DL, pero no tan a menudo como *DCDC2* y
KIAA0319. Sin embargo, con el panel de SNP denso descrito en el presente documento, fue posible observar
 asociaciones con otros elementos de *DYX2*, incluidos *FAM65B* y *CMAHP*.

Otra posible explicación para el número de genes de *DYX2* observados que se asocian a los fenotipos en este
 estudio es el desequilibrio de enlace DE dentro del locus *DYX2*. De hecho, es probable que el DE explique el grupo
 45 de asociaciones alrededor de *KIAA0319*, *TDP2*, *ACOT13* y *C6orf62*. Como se describe aquí, dos bloques con DE
 principales abarcan *KIAA0319*: uno abarca la mitad 3' del gen, mientras que el otro abarca la región 5' de *KIAA0319*,
 así como *ACOT13*, *TDP2* y parte de *C6orf62*. Casi todas las asociaciones en este estudio se localizan en este
 bloque con DE 5', que también contiene el haplotipo de riesgo de DL *KIAA0319* reportado anteriormente. Debido a
 esta estructura con DE, es imposible determinar si las asociaciones en esta región son independientes o si capturan
 50 la misma variante funcional. La última posibilidad se considera la más probable, y se cree que las asociaciones en
 esta región probablemente estén marcando la misma variante causal capturada por el haplotipo de riesgo de DL
KIAA0319. El estudio funcional de esta región, particularmente de los genes menos estudiados, *TDP2*, *ACOT13* y
C6orf62, será útil para determinar si estas asociaciones son independientes o no.

En contraste, los marcadores dentro o cerca de *FAM65B* y *CMAHP* parecen estar capturando distintas señales de
 asociación de dos bloques con DE diferentes (datos no mostrados). El SNP rs9348646, que mostró asociación con
 IC verbal, se localiza dentro de un intrón de *FAM65B* en un bloque con DE, mientras que rs1562422, que mostró
 asociación con DL moderada, se localiza en un bloque con DE separado. Si bien rs1562422 es un marcador
 intergénico ubicado físicamente entre *FAM65B* y *CMAHP*, muestra un DE fuerte con marcadores en *CMAHP* (datos
 no mostrados). Los patrones de DE dentro del locus *DYX2* sugieren que las asociaciones en *KIAA0319*, *TDP2*,

ACOT13 y *C6orf62* están marcando la misma variante causal, mientras que rs9348646 en *FAM65B* y rs1562422 son independientes.

Los otros genes de *DYX2*, incluidos *FAM65B* y *CMAHP*, se han estudiado mucho menos que los genes de riesgo *DCDC2* y *KIAA0319*. Poco se sabe acerca de *FAM65B* en términos de función biológica; sin embargo, existe evidencia de que *FAM65B* puede influir en la migración en los linfocitos T (Rougerie *et al.* 2013). Los modelos animales de *DCDC2* y *KIAA0319* han implicado a estos genes en procesos migratorios, aunque en un contexto neuronal. *CMAHP*, que codifica una enzima clave en la síntesis de los ácidos siálicos Neu5Ac y Neu5Gc en otros mamíferos, se convirtió en un pseudogén en humanos por una microdelección inactivadora y la subsiguiente fijación del alelo inactivo en poblaciones humanas tempranas (Chou *et al.* 1998). Aunque *ACOT13* parece estar marcando la variación dentro de *KIAA0319*, los estudios funcionales preliminares de *ACOT13* resultan intrigantes. *ACOT13* se asoció recientemente a una activación asimétrica más baja del surco temporal superior posterior durante las tareas de lectura y fonología (Pinel *et al.* 2012). El producto proteico codificado por *ACOT13* ha sido colocalizado con beta-tubulina en microtúbulos; se propone que la unión a los microtúbulos es importante para la DL, ya que *DCDC2* contiene dos dominios de doblecortina que se cree que se unen a los microtúbulos (Cheng *et al.* 2006).

Los genes y elementos reguladores dentro del locus *DYX2* pueden contribuir interactivamente a los dominios de la lectura y el lenguaje, como se ve con la relación no aditiva entre las variantes reguladoras putativas en *DCDC2* y *KIAA0319* (Ejemplo 1; Powers *et al.* 2013, Ludwig *et al.* 2008). Se ha demostrado que estas variantes de riesgo influyen en la expresión génica y que interactúan entre sí para influir sustancialmente en el rendimiento en las tareas de la lectura y el lenguaje. Una red reguladora compleja, donde los elementos reguladores interactúan y corregulan otros genes y elementos de *DYX2*, podría contribuir a la lectura, el lenguaje y la cognición. Si es así, es probable que el elemento READ1 en *DCDC2* y la variante causal marcada por el haplotipo de riesgo de *KIAA0319* tengan los efectos más fuertes sobre la expresión génica y el fenotipo neurocognitivo definitivo. Respalda esta idea el hecho de que muchos de los aciertos de asociación descritos en el presente documento, tanto basados en marcadores únicos como en haplotipos, y con los tres fenotipos, se localizan en el mismo bloque con DE que el haplotipo de riesgo de *KIAA0319*. Este resultado, junto con la asociación del haplotipo de riesgo de *KIAA0319* con expresión reducida de *KIAA0319* y su interacción sinérgica con un elemento regulador en un intrón de *DCDC2*, indica la presencia de al menos una variante reguladora en esta región que influye en la expresión de *KIAA0319*. Las ubicaciones de los únicos otros aciertos independientes en el locus (aparte de READ1 en *DCDC2*), un intrón de *FAM65B* y corriente abajo de un pseudogén, pueden sugerir la presencia de regiones reguladoras adicionales que influyen en la expresión génica. Según el trabajo descrito en el presente documento, *DCDC2* y *KIAA0319* son los principales genes efectores responsables de la influencia de *DYX2* en el riesgo de DL y TL, y la alteración de los niveles o patrones de expresión génica es el mecanismo por el cual se ejerce este efecto.

En el trabajo descrito en el presente documento, se enfatizó la replicación de asociaciones genéticas en cohortes independientes, en lugar de la confianza en correcciones estadísticas para pruebas múltiples, para la validación de las asociaciones en la cohorte de descubrimiento de ALSPAC. Las replications de la asociación genética con los rasgos neurocognitivos de interés, en particular en las diversas cohortes de esta investigación, proporcionan pruebas sólidas de que los resultados de este estudio no se deben a un error de Tipo I. También se reportan valores *p* no corregidos y un umbral estadístico que corrige 195 marcadores genéticos (umbral de 0,000256) para presentar el contexto de los hallazgos en términos de fuerza y confiabilidad. No obstante, las tres cohortes de replicación no eran idénticas y tenían diferencias inherentes entre sí y con respecto a la cohorte de descubrimiento, que pudieron haber evitado la replicación. Estas diferencias incluyen: (1) el trastorno para el que se seleccionó cada cohorte (DL vs. TL vs. no seleccionado), (2) la gravedad de la definición del caso y el reclutamiento, y (3) el país de reclutamiento (Reino Unido vs. EE. UU. vs. Italia) y el idioma hablado (inglés vs. italiano). Por ejemplo, el uso de un idioma regular como el italiano en lugar de un idioma irregular como el inglés puede haber permitido detectar más fácilmente las verdaderas deficiencias en la lectura y el lenguaje. Este problema probablemente se haya evitado en la cohorte de descubrimiento de ALSPAC debido al gran tamaño de la muestra. La observación de múltiples asociaciones replicadas en todo el locus *DYX2* descrito en el presente documento aumenta la confianza en estos resultados.

En resumen, los análisis indican cuatro señales de asociación para DL, TL y Coeficiente verbal en el locus *DYX2*: *DCDC2*, *KIAA0319*, *FAM65B* y el pseudogén *CMAHP*. Los resultados de asociación dentro de las áreas de *DCDC2* y *KIAA0319* (incluyendo *TDP2*, *ACOT13* y *C6orf62*) están en DE con dos variantes de riesgo reportadas previamente: el elemento regulador READ1 en *DCDC2* y el haplotipo de riesgo *KIAA0319* en *KIAA0319* y *TDP2*. Estos resultados apuntan fuertemente a la variación en la expresión del gen *KIAA0319* como mediadora del efecto de *DYX2* en los fenotipos de la lectura y el lenguaje.

Referencias

- Barrett JC, *et al.* (2005) *Bioinformatics* 21:263-265.
- Beaver KM, *et al.* (2010) *J Neural Transm* 117(7):827-30.
- Bishop DV, *et al.* (2008) *Genes Brain Behav* 7(3):365-72.
- Bishop DV, *et al.* (1996) *J Child Physiol Psychiatry* 37:391-403.

- Boyd A, et al. (2012) *Int J Epidemiol* 42(1):111-27.
- Cardon LR, et al. (1994) *Science* 266(5183):276-9.
- Catts HW, et al. (2005) *J Speech Lang Hear Res* 48(6):1378-96.
- Cheng Z, et al. (2006) *Biochem Biophys Res Commun* 350(4):850-3.
- 5 Chou HH, et al. (1998) *Proc Natl Acad Sci USA* 95(20):11751-6.
- Cope N, et al. (2005) *Am J Hum Genet* 76(4):581-91.
- Couto JM, et al. (2010) *Am J Med Genet B Neuropsychiatr Genet* 153B(2):447-62.
- Deffenbacher KE, et al. (2004) *Hum Genet* 115(2):128-38.
- Dennis MY, et al. (2009) *PLoS Genet* 5(3):e1000436.
- 10 Eicher JD y Gruen JR. (2013) *Mol Genet Metab*, doi: 10.1016/j.ymgme.2013.07.001.
- Eicher JD, et al. (2013a) *PLoS One* 8(5):e63762.
- Eicher JD, et al. (2013b) *Genes Brain Behav*, en prensa.
- Elbert A, et al. (2011) *Behav Genet* 41(1):77-89.
- Francks C, et al. (2004) *Am J Hum Genet* 75(6):1046-1058.
- 15 Gathercole S y Baddeley AD. (1990) *Journal of Memory and Language* 29:336-360.
- Gathercole SE y Baddeley AD. (1993) Lawrence Erlbaum, Mahwah, NJ.
- Gathercole SE y Baddeley AD. (1996) The Psychological Corporation, Londres.
- Gayán J, et al. (1999) *Am J Hum Genet* 64(1):157-64.
- Golding J, et al. (2001) *Paediatr Perinat Epidemiol* 15(1):74-87.
- 20 Graham SA y Fisher SE. (2013) *Curr Opin Neurobiol* 23(1):43-51.
- Han B, et al. (2008) *Ann Hum Genet* 72(Pt 6):834-847
- Harold D, et al. (2006) *Mol Psychiatry* 11(12):1085-1091.
- Kaminen N, et al. (2003) *J Med Genet* 40(5):340-5.
- Kaplan DE, et al. (2002) *Am J Hum Genet* 70(5):1287-98.
- 25 Landi N, et al. (2013) *Dev Sci* 16(1):13-23.
- Lind PA, et al. (2010) *Eur J Hum Genet* 18(6):668-73.
- Luciano M, et al. (2007) *Biol Psychiatry* 62:811-817.
- Ludwig KU, et al. (2008) *J Neural Transm* 115(11):1587-9.
- Marino C, et al. (2012) *Psychiatr Genet* 22(1):25-30.
- 30 Meng H, et al. (2011) *Behav Genet* 41(1):58-66.
- Meng H, et al. (2005) *Proc Natl Acad Sci USA* 102:17053-17058.
- Newbury DF, et al. (2010) *Genome Med* 2(1):6.
- Newbury DF, et al. (2011) *Behav Genet* 41(1):90-104.
- Newbury DF, et al. (2009) *Am J Hum Genet* 85(2):264-72.
- 35 Nunes T, et al. (2003) *Scientific Studies of Reading* 7(3):289-307.
- Paracchini S, et al. (2008). *Am J Psychiatry* 165(12):1576-1584.
- Paracchini S, et al. (2006) *Hum Mol Genet* 15(10):1659-1666.

Pennington BF. (2006) *Cognition* 101(2):385-413.

Pennington BF y Bishop DV (2009) *Annual Review of Psychology* 60:283-306.

Peter B, et al. (2011) *J Neurodev Disord* 3(1):39-49.

Pinel P, et al. (2012) *J Neurosci* 32(3):817-25.

5 Plomin R, et al. (2004) *Mol Psychiatry* 9(6):582-6.

Powers NR, et al. (2013) *Am J Hum Genet* 93(1):19-28.

Purcell S, et al. (2007) *Am J Hum Genet* 81(3):559-575.

Rosner J y Simon DP. (1971) *Journal of Learning Disabilities* 4(384):40-48.

Rougerie P, et al. (2013) *J Immunol* 190(2):748-55.

10 Rust J, et al. (1993) Psychological Corporation, Sidcup, Reino Unido.

Scerri TS, et al. (2011) *Biol Psychiatry* 70:237-245.

Schumacher J, et al. (2006) *Am J Hum Genet* 78(1):52-62.

Trzaskowski M, et al. (2013) *Behav Genet* 43(4):267-73.

Wechsler D. (1996) Psychological Corporation, Londres, Reino Unido.

15 Wechsler D, et al. (1992) Psychological Corporation, Sidcup, Reino Unido.

Weismer SE, et al. (2000) *J Speech Lang Hear Res* 43(4):865-78.

Wilcke A, et al. (2011) *Eur J Hum Genet* 20(2):224-9.

Wilcke A, et al. (2009) *Ann Dyslexia* 59(1):1-11.

Wise JC, et al. (2007) *J Speech Lang Hear Res* 50(4):1093-9.

20 Wong PC, et al. (2013) *PLoS One* 8(5):e64983.

Viding E, et al. (2004) *J Child Psychol Psychiatry* 45(2):315-25.

Zhong R, et al. (2013) *Mol Neurobiol* 47(1):435-42.

Zou L, et al. (2012) *Am J Med Genet B Neuropsychiatr Genet* 159B(8):970-6.

Leyendas de las figuras

- 25 Figura 1: Esquema de los genes dentro del locus DYX2 en el cromosoma 6p21.3. Los genes en azul, *DCDC2* y *KIAA0319*, tienen asociaciones replicadas con fenotipos del lenguaje verbal y escrito, a saber, de DL y TL. Las regiones en rojo marcan dos variantes funcionales, *READ1* en *DCDC2* y un haplotipo de riesgo con marcadores en *KIAA0319* y *TDP2*, que se han asociado funcionalmente a DL y TL, utilizando modelos animales y técnicas moleculares.
- 30 Figura 4: Un esquema actualizado de genes en nuestro estudio con marcadores que muestran asociaciones replicadas a DL, TL y/o CI. La lista de estos genes (mostrados en azul) se ha ampliado a siete (*DCDC2*, *KIAA0319*, *TDP2*, *ACOT13*, *C6orf62*, *FAM65B* y *CMAHP*), aunque el desequilibrio de enlace pueda dar cuenta de múltiples asociaciones (en particular para *KIAA0319*, *TDP2*, *ACOT13* y *C6orf62*).

Tablas del Ejemplo 3

35 Tabla 1a: Mediciones del fenotipo de ALSPAC

Medición	Dominio
Eliminación de fonemas (PD) Edad 7 años	Lectura (DL)
Lectura de palabras sueltas (SWR7) Edad 7 años	Lectura (DL)

ES 2 718 343 T3

Medición	Dominio
Lectura de pseudopalabras sueltas (SNR) Edad 7 años	Lectura (DL)
Lectura de palabras sueltas (SWR9) Edad 9 años	Lectura (DL)
Dimensiones Objetivas del Lenguaje de Wechsler (WOLD) Comprensión verbal Edad 8 años	Idioma (TL)
Tarea de repetición de pseudopalabras (NWR) Edad 8 años	Idioma (TL)
Escala de inteligencia de Wechsler para niños (WISC) CI total (CIT) Edad 8 años	Cognición (CI)
Escala de inteligencia de Wechsler para niños (WISC) CI verbal (CIV) Edad 8 años	Cognición (CI)
Escala de inteligencia de Wechsler para niños (WISC) CI de rendimiento (CIR) Edad 8 años	Cognición (CI)

Tabla 1b: Definiciones de fenotipos para análisis de ALSPAC

Lectura (DL)	Definición de fenotipo
DL grave	2 desviaciones estándar por debajo de la media de la muestra en la tarea de eliminación de fonemas
DL moderada	1 desviación estándar por debajo de la media de la muestra en las tareas SWR7, SNR y SWR
Idioma (TL)	Definición de fenotipo
TL grave	2 desviaciones estándar por debajo de la media de la muestra en las tareas WOLD y/o NWR
WOLD moderado	1,5 desviaciones estándar por debajo de la media de la muestra en la tarea WOLD
NWR moderada	1,5 desviaciones estándar por debajo de la media de la muestra en la tarea NWR
Cognición (CI)	Definición de fenotipo
CI total	Rendimiento cuantitativo en la tarea de CI Total de WISC
CI verbal	Desempeño cuantitativo en la tarea de CI Verbal de WISC
CI de rendimiento	Rendimiento cuantitativo en la tarea de CI de Rendimiento de WISC

Tabla 2: Cohortes de replicación

	TL Iowa	DL Colorado	DL Italia
Número de sujetos	428	1201	878
Numero de familias	N/A	293	304
Trastorno	TL	DL	DL

ES 2 718 343 T3

	TL Iowa	DL Colorado	DL Italia
Tipo de cohorte	Control de caso	Basada en la familia	Basada en la familia
Análisis	SVS	TDT (PLINK)	TDT (PLINK)
Asociación	Estado de control de caso	Estado de control de caso y	Estado de control de caso
Condicionado en:		Puntuación discriminante	
Estado del caso determinado en:	Puntuación compuesta en mediciones del lenguaje		Velocidad o precisión en la tarea de lectura de palabras sueltas o texto

Tabla 3: Asociaciones genéticas de marcadores únicos con varias definiciones de casos y controles de DL y TL.

Fenotipo	Marcador	Gen	Ubicación de BP	Modelo	OR (IC del 95 %)	Valor p
DL grave	rs2294691	<i>TDP2</i>	24652843	Alélico	2,0 (1,3-2,9)	0,00050
DL grave	rs2294691	<i>TDP2</i>	24652843	Aditivo	1,9 (1,3-2,8)	0,00053
DL grave	rs2294691	<i>TDP2</i>	24652843	Dominante	2,3 (1,5-3,7)	0,00018*
DL grave	rs10456309	<i>KIAA0319</i>	24589562	Recesivo	10,5 (2,2-49,5)	0,00020*
DL moderada	rs1562422	<i>CMAHP</i>	25044577	Dominante	1,7 (1,2-2,2)	0,00081
TL grave	rs807694	<i>DCDC2</i>	24303383	Aditivo	1,8 (1,3-2,5)	0,00057
TL grave	rs807694	<i>DCDC2</i>	24303383	Alélico	1,8 (1,3-2,5)	0,00050
TL grave	rs807694	<i>DCDC2</i>	24303383	Dominante	1,9 (1,3-2,7)	0,00062
WOLD moderado	rs3756814	<i>C6orf62</i>	24705835	Aditivo	0,7 (0,6-0,9)	0,00039
WOLD moderado	rs3756814	<i>C6orf62</i>	24705835	Alélico	0,7 (0,6-0,9)	0,00047
WOLD moderado	rs3777663	<i>ACOT13</i>	24700235	Aditivo	0,6 (0,5-0,8)	0,00039
WOLD moderado	rs3777663	<i>ACOT13</i>	24700235	Alélico	0,7 (0,5-0,8)	0,00041

*La asociación genética sobrevive a la corrección por múltiples pruebas.

5 Tabla 4: Asociaciones genéticas basadas en marcadores únicos (Tabla 4a) y haplotipos (Tabla 4b) con medición cuantitativa de la cognición.

Tabla 4a: Asociaciones genéticas de marcadores únicos con la cognición

Fenotipo	Marcador	Gen	Ubicación de BP	Modelo	Pendiente	Valor p
----------	----------	-----	-----------------	--------	-----------	---------

ES 2 718 343 T3

Fenotipo	Marcador	Gen	Ubicación de BP	Modelo	Pendiente	Valor p
CI verbal	rs9295626	<i>KIAA0319</i>	24587339	Alélico	1,40	0,00041
CI verbal	rs9295626	<i>KIAA0319</i>	24587339	Aditivo	1,39	0,00043
CI verbal	rs7763790	<i>KIAA0319</i>	24615063	Alélico	-1,40	0,00045
CI verbal	rs7763790	<i>KIAA0319</i>	24615063	Aditivo	-1,38	0,00048
CI verbal	rs6935076	<i>KIAA0319</i>	24644322	Alélico	1,16	0,00049
CI verbal	rs6935076	<i>KIAA0319</i>	24644322	Aditivo	1,15	0,00052
CI verbal	rs9348646	<i>FAM65B</i>	24052526	Alélico	-1,14	0,00066
CI verbal	rs9348646	<i>FAM65B</i>	24052526	Aditivo	-1,14	0,00066
CI total	rs2328791	N/A	23736848	Alélico	-1,21	0,00066
CI total	rs2328791	N/A	23736848	Aditivo	-1,18	0,00075
CI total	rs2328791	N/A	23736848	Recesivo	-3,36	0,00042

Tabla 4b: Asociaciones genéticas basadas en haplotipos con cognición

Marcadores	Haplotipo	Gen	Ubicación de BP	Pendiente	Valor p
rs2817201, rs9295626	AT	<i>KIAA0319</i>	24585214, 24587339	1,42	0,000378
rs10456309, rs4576240, rs17307478, rs9356939, rs7763790, rs6456621	GGTCAC	<i>KIAA0319</i>	24589562, 24596478, 24605024, 24613354, 24615063, 24618511	-1,40	0,000569
rs6456624, rs6935076, rs2038137, rs3756821, rs1883593, rs3212236	AGATA	<i>KIAA0319</i>	24639223, 24644322, 24645943, 24646821, 24647191, 24648455	1,81	0,0000145*
rs3777663, rs3756814, rs6931809, rs6916186, rs6933328, rs17491647	TGTGGA	<i>ACOT13/C6orf62</i>	24700235, 24705835, 24706770, 24708523, 24710920, 24713723	-1,56	0,000742

*La asociación genética sobrevive a la corrección por múltiples pruebas.

Tabla 5: Replicación de asociaciones genéticas en las cohortes de Iowa, Italia y Colorado.

Marcador	Gen	Control de caso de Iowa		Control de caso de Italia		Control de caso de Colorado		Puntuación discriminante de Colorado	
		OR	p	OR	p	OR	p	Pendiente	p
rs2328791	N/A	1,0	0,813	1,0	1,000	0,9	0,646	0,087	0,447

ES 2 718 343 T3

Marcador	Gen	Control de caso de Iowa		Control de caso de Italia		Control de caso de Colorado		Puntuación discriminante de Colorado	
		OR	p	OR	p	OR	p	Pendiente	p
rs33914824^a	DCDC2	2,2	0,034	0,9	0,768	1,1	0,847	0,023	0,934
rs807694^a	DCDC2	1,9	0,028	0,9	0,786	0,9	0,853	-0,025	0,919
rs707864^a	DCDC2	1,6	0,017	1,0	0,840	1,2	0,446	-0,246	0,101
rs10456301 ^a	DCDC2	0,9	0,553	1,1	0,811	1,5	0,289	0,221	0,162
rs16889066 ^a	DCDC2	1,2	0,517	1,0	0,884	1,2	0,622	-0,304	0,150
rs9379651 ^a	DCDC2	1,1	0,602	1,3	0,225	0,6	0,059	0,205	0,141
rs2817201	KIAA0319	1,1	0,733	1,2	0,129	1,0	1,000	0,034	0,787
rs9295626	KIAA0319	1,1	0,579	0,6	0,0055	1,0	0,823	-0,158	0,169
rs10456309	KIAA0319	0,5	0,073	0,7	0,189	0,4	0,206	0,628	0,00845
rs4576240	KIAA0319	1,1	0,825	1,9	0,0027	1,1	0,862	-0,052	0,754
rs17307478	KIAA0319	1,0	0,996	1,3	0,292	0,8	0,555	0,039	0,803
rs9356939	KIAA0319	4,0	0,018	0,8	0,069	1,3	0,151	-0,116	0,254
rs7763790	KIAA0319	1,0	0,831	1,1	0,627	1,4	0,163	0,014	0,910
rs6456621	KIAA0319	2,2	0,019	1,6	0,405	1,8	0,366	-0,458	0,104
rs3756821	KIAA0319	1,2	0,278	1,0	0,842	1,2	0,327	-0,033	0,734
rs1883593	KIAA0319	1,3	0,169	1,6	0,0052	1,3	0,239	-0,108	0,395
rs3212236	KIAA0319	1,0	0,883	1,1	0,496	0,9	0,745	-0,124	0,319
rs2294691	TDP2	1,1	0,779	1,9	0,0578	1,4	0,491	-0,290	0,247
rs3777663	ACOT13	0,7	0,016	0,6	0,0052	1,0	0,908	0,101	0,345
rs3756814	C6orf62	0,7	0,005	0,7	0,023	0,9	0,600	-0,003	0,980
rs6931809	C6orf62	1,4	0,023	1,4	0,017	1,2	0,491	-0,096	0,382
rs6916186	C6orf62	0,9	0,757	1,2	0,413	1,2	0,547	0,112	0,490
rs6933328	C6orf62	0,9	0,612	0,9	0,613	1,0	0,827	0,215	0,0437
rs17491647	C6orf62	0,8	0,155	0,7	0,104	1,0	0,901	0,042	0,709
rs9348646	FAM65B	0,9	0,358	1,1	0,535	1,4	0,144	-0,415	0,00021

Marcador	Gen	Control de caso de Iowa		Control de caso de Italia		Control de caso de Colorado		Puntuación discriminante de Colorado	
		OR	p	OR	p	OR	p	Pendiente	p
rs1562422	CMAHP	1,0	0,793	1,0	0,796	0,6	0,093	-0,030	0,840

^aEstos marcadores son parte del haplotipo de riesgo de seis marcadores en *DCDC2* analizado en detalle en Powers et al. 2013.

Tabla 6: Correlaciones de fenotipos en la cohorte de ALSPAC*

	NWR	WOLD	SWR7	SWR9	SNR	PD	CIT	CIV	CIR
NWR	1								
WOLD	0,214	1							
SWR7	0,403	0,259	1						
SWR9	0,351	0,202	0,722	1					
SNR	0,306	0,149	0,660	0,708	1				
PD	0,362	0,165	0,688	0,550	0,538	1			
CIT	0,324	0,386	0,500	0,387	0,343	0,406	1		
CIV	0,346	0,424	0,536	0,421	0,421	0,426	0,871	1	
CIR	0,192	0,216	0,292	0,218	0,218	0,246	0,819	0,435	1

*Todas las correlaciones tuvieron $p < 0,05$.

Ejemplo 4

- 5 Estudio de Asociación del Genoma Completo de Componentes Compartidos de la Discapacidad Lectora y el Trastorno del Lenguaje; *ZNF385D* influye en los trastornos de la lectura y el lenguaje.

Introducción

Tanto DL como TL son rasgos complejos que se comanifiestan con frecuencia, lo que lleva a la hipótesis de que estos trastornos comparten una etiología genética. Para probar esto, se realizó un estudio de asociación del genoma completo (GWAS) en individuos afectados con DL y TL en el Estudio Longitudinal Avon de Padres e Hijos. Las asociaciones más fuertes se vieron con marcadores en *ZNF385D* ($OR = 1,81, p = 5,45 \times 10^{-7}$) y *COL4A2* ($OR = 1,71, p = 7,59 \times 10^{-7}$). Los marcadores dentro de *NDST4* mostraron las asociaciones más fuertes con TL individualmente ($OR = 1,827, p = 1,40 \times 10^{-7}$). La asociación de *ZNF385D* se replicó utilizando mediciones del vocabulario receptivo en el estudio de Genética Neurocognitiva Pediátrica en Imágenes ($p = 0,00245$). Luego se utilizaron los datos de volumen de tractos de fibras del tensor de difusión en 16 tractos de fibra para examinar las implicaciones de los marcadores replicados. *ZNF385D* fue un predictor de los volúmenes generales de tractos de fibra en ambos hemisferios, así como del volumen global del cerebro. En este ejemplo, se presenta evidencia para *ZNF385D* como un gen de riesgo para DL y TL. La implicación del factor de transcripción de *ZNF385D* en DL y TL subraya la importancia de la regulación transcripcional en el desarrollo de rasgos neurocognitivos de orden superior.

20 Métodos

ALSPAC.

El equipo de ALSPAC llevó a cabo el reclutamiento de sujetos y la recopilación de fenotipos y datos genéticos para la cohorte de ALSPAC. El ALSPAC es una cohorte de nacimientos prospectiva, basada en la población, con sede en la región de Avon en el Reino Unido. Se compone principalmente de niños de ascendencia del norte de Europa, nacidos en 1991 y 1992. Los niños fueron reclutados antes del nacimiento; el reclutamiento de sus madres embarazadas dio lugar a un total de 15.458 fetos, de los cuales 14.701 estaban vivos al año de edad. Los detalles sobre los participantes, el reclutamiento y las metodologías de estudio se describen en detalle en otra parte (<http://www.bristol.ac.uk/alspac>) (Boyd *et al.*, 2012; Golding *et al.*, 2001). Los niños de ALSPAC han sido ampliamente fenotipados desde antes del nacimiento hasta la edad adulta temprana. La aprobación ética se obtuvo del Comité de Ética y Derecho de ALSPAC, los comités locales de ética en investigaciones del Reino Unido y el Comité de Investigación con Seres Humanos de Yale.

Mediciones de la lectura y el lenguaje.

Las mediciones de la lectura, el lenguaje y la cognición utilizadas para este estudio se recopilaron a las edades de 7, 8 y 9 años. Los sujetos con $CI \leq 75$ en el CI Total de la Escala de Inteligencia de Wechsler para Niños (WISC-III), realizado a los 8 años de edad, fueron excluidos de los análisis presentados (Wechsler *et al.* 1992). Las mediciones de la lectura en el ALSPAC incluyen una tarea de eliminación de fonemas a los 7 años de edad, la lectura de palabras sueltas a los 7 y 9 años de edad, la lectura de seudopalabras a los 9 años de edad, y la comprensión lectora de un texto a los 9 años de edad. La tarea de eliminación de fonemas evalúa el reconocimiento de los fonemas, considerado ampliamente como un déficit central tanto en DL como en TL (Pennington, 2006; Pennington y Bishop, 2009). Para la tarea de eliminación de fonemas, también conocida como Prueba de Análisis Auditivo, el niño escucha una palabra hablada en voz alta y luego se le pide que elimine un fonema específico de esa palabra para formar una nueva palabra (Rosner y Simon, 1971). La lectura de palabras sueltas se evaluó a los 7 años de edad utilizando la subprueba de lectura de las Dimensiones Objetivas de la Lectura de Wechsler (WORD). A los 9 años de edad, se evaluó la lectura de palabras y seudopalabras sueltas pidiéndole al niño que leyera diez palabras reales y diez seudopalabras en voz alta de un subconjunto de una lista más grande de palabras y seudopalabras tomadas de la investigación realizada por Terezinha Nunes y colegas (Rust *et al.*, 1993). Las puntuaciones de comprensión lectora se determinaron a los 9 años de edad, utilizando el Análisis de la Capacidad Lectora de Neale (NARA-II) (Neale 1997). Se recopilaron dos mediciones del lenguaje adicionales, las tareas de repetición de seudopalabras y de comprensión verbal, durante entrevistas clínicas a los 8 años de edad. Se utilizó una adaptación de la Tarea de Repetición de Seudopalabras (NWR), en la que los sujetos repitieron grabaciones de seudopalabras, para evaluar el procesamiento y la memoria fonológica a corto plazo (Gathercole y Baddeley, 1996). Los niños también completaron la tarea de comprensión verbal de las Dimensiones Objetivas del Lenguaje de Wechsler (WOLD), en la que respondieron a preguntas sobre un párrafo leído en voz alta por un examinador que describía una imagen presentada (Wechsler 1996). Se calcularon las puntuaciones Z para cada sujeto en cada medición individual.

Definiciones de casos.

El solicitante se propuso capturar de manera persistente el bajo desempeño en varios dominios de la lectura y el lenguaje verbal como casos de DL y TL en las definiciones de casos (Tabla 1). Por lo tanto, se definió que los casos de DL tenían una puntuación z menor o igual a -1 en al menos 3 de las 5 tareas siguientes: lectura de palabras sueltas a los 7 años de edad, eliminación de fonemas a los 7 años de edad, lectura de palabras sueltas a los 9 años de edad, lectura de seudopalabras a los 9 años de edad y comprensión lectora a los 9 años de edad. Hubo 527 sujetos definidos como casos de DL. Se definió que los casos de TL tenían una puntuación z menor o igual a -1 en al menos 2 de las 3 tareas siguientes: eliminación de fonemas a los 7 años de edad, comprensión verbal a los 8 años de edad y repetición de seudopalabras a los 8 años de edad. Hubo 337 sujetos definidos como casos de TL. Como el reconocimiento de fonemas es importante tanto en DL como en TL, se incluyó en la definición de casos, tanto para DL como TL, con el fin de reflejar la presentación clínica. Hubo 174 individuos afectados con DL y TL, con una proporción de hombres y mujeres de 1,7:1. En la caracterización adicional de las asociaciones observadas, se crearon subconjuntos de casos sin comorbilidad. Hubo 163 casos de TL, excluyendo aquellos con DL concomitante, y 353 casos de DL, excluyendo aquellos con TL concomitante. Para todos los análisis, los controles se definieron como sujetos de ALSPAC de ascendencia europea que realizaron todas las evaluaciones neuroconductuales necesarias, pero no cumplieron con los criterios para el estado del caso.

Genotipado y análisis.

Los sujetos fueron genotipados en matrices de cuentas Illumina HumanHap 550 (San Diego, CA). Los sujetos eran excluidos si el porcentaje de genotipos faltantes era superior al 2 % ($n = 6$). Para evitar una posible estratificación de la población, solo se incluyeron sujetos de ascendencia europea. En el análisis primario de individuos con DL y TL, hubo 174 casos y 4117 controles. Hubo un total de 500.527 SNP genotipados antes de la evaluación de calidad y el control de calidad. Los marcadores se eliminaban si el equilibrio de Hardy-Weinberg $p \leq 0,0001$ ($n = 93$) o si la falta era mayor que 10 % ($n = 19$). Todos los marcadores tenían una frecuencia de alelo menor mayor que 0,01. Todos los análisis genéticos se realizaron mediante regresión logística en PLINK v1.07 (Purcell *et al.*, 2007). Para corregir las pruebas múltiples, se estableció un umbral corregido de Bonferroni de $\alpha = 1,00 \times 10^{-7} = 0,05/500.000$ marcadores probados.

Después de los análisis iniciales que examinaron los casos con DL y TL, se examinaron las definiciones de los casos de DL y TL de forma individual (por ejemplo, TL excluyendo aquellos con DL concomitante, y DL excluyendo aquellos con TL concomitante). Estos análisis se llevaron a cabo para determinar si un solo trastorno (DL o TL) impulsaba señales de asociación en el análisis de DL y TL concomitantes. También se examinaron las asociaciones de marcadores dentro de varios genes de riesgo de DL y/o TL previamente identificados, incluidos los reportados recientemente en Luciano *et al.*, para presentar sus resultados con estas definiciones fenotípicas. Estos genes incluían: *ABCC13*, *ATP2C2*, *BC0307918*, *CMIP*, *CNTNAP2*, *DAZAP1*, *DCDC2*, *DYX1C1*, *FOXP2*, *KIAA0319*, *KIAA0319L*, *PRKCH*, *ROBO1* y *TDP2*.

Se realizaron análisis genéticos en cada fenotipo (DL y TL concomitantes, así como DL y TL individualmente) utilizando el programa VEGAS, similar al del estudio de Luciano *et al.* (Liu *et al.*, 2010; Luciano *et al.*, 2013). Para corregir las pruebas múltiples, se estableció un umbral corregido de Bonferroni de $\alpha = 2,84 \times 10^{-6} = 0,05/17,610$ genes probados.

Análisis de replicación de PING.

Los análisis de replicación se llevaron a cabo en el estudio PING. Los detalles sobre los métodos de reclutamiento, de constatación, neuroconductuales, genéticos y neuroimagenológicos, y la adquisición de datos en el estudio PING, se describen en detalle en otra parte, pero se resumen brevemente a continuación (Akshoomoff *et al.*, 2013, Brown *et al.*, 2012; Fjell *et al.*, 2012; Walhovd *et al.*, 2012). El estudio PING es una cohorte transversal de niños y adolescentes con un desarrollo típico entre las edades de 3 y 20 años. Los sujetos fueron evaluados en busca de antecedentes de trastornos importantes del desarrollo, psiquiátricos y/o neurológicos, lesión cerebral o afecciones médicas que afectan el desarrollo. Sin embargo, los sujetos no fueron excluidos debido a problemas de aprendizaje como DL y TL. Los programas de protección para investigaciones con seres humanos y las juntas de revisión institucional en las 10 instituciones (Weill Cornell Medical College, Universidad de California en Davis, Universidad de Hawái, Instituto Kennedy Krieger, Hospital General de Massachusetts, Universidad de California en Los Ángeles, Universidad de California en San Diego, Facultad de Medicina de la Universidad de Massachusetts, Universidad del Sur de California y Universidad de Yale) que participaron en el estudio PING aprobaron todos los procedimientos experimentales y de consentimiento. Para individuos menores de 18 años de edad, se obtuvo el consentimiento informado de los padres y el asentimiento del menor (para las personas de 7 a 17 años de edad). Todos los participantes mayores de 18 años de edad dieron su consentimiento informado por escrito.

Los sujetos realizaron la versión del estudio validada de NIH Toolbox Cognition Battery, en la que se llevaron a cabo dos tareas relacionadas con el lenguaje y la lectura: la Prueba de Reconocimiento de Lectura Oral y la Prueba de Vocabulario con Imágenes (Akshoomoff *et al.*, 2013; Weintraub *et al.*, 2013). En la Prueba de Reconocimiento de Lectura Oral, una palabra o letra se presenta en la pantalla de la computadora y se le pide al participante que la lea en voz alta. Las respuestas son registradas como correctas o incorrectas por el examinador, quien ve las pronunciaciones aceptadas en una pantalla de computadora aparte. La Prueba de Vocabulario con Imágenes es una medición del vocabulario receptivo y se administra en un formato adaptativo computarizado. Al participante se le presenta una grabación auditiva de una palabra y cuatro imágenes en la pantalla de la computadora; la tarea es tocar la imagen que mejor represente el significado de la palabra.

Los sujetos fueron genotipados en el Illumina Human660W-Quad BeadChip (San Diego, CA), con marcadores utilizados para análisis de replicación que pasaron los filtros de control de calidad (tasa de llamada de muestra >98 %, tasa de llamada de SNP >95 %, frecuencia de alelos menores >5 %). Se construyó un panel de referencia como se describe en otra parte (Brown *et al.*, 2012; Fjell *et al.*, 2012; Walhovd *et al.*, 2012). Para evaluar las proporciones de ascendencia y mezcla en los participantes de PING, se utilizó un enfoque de agrupación supervisado implementado en el software ADMIXTURE (Alexander *et al.*, 2009), y se agruparon los datos de los participantes en seis grupos correspondientes a seis poblaciones continentales principales: Africano, asiático central, asiático oriental, europeo, nativo americano y oceánico. La implementación de proporciones de ascendencia y mezcla en los sujetos de PING se describe en detalle en otra parte (Brown *et al.*, 2012; Fjell *et al.*, 2012; Walhovd *et al.*, 2012). Para evitar una posible estratificación de la población, solo se incluyeron en el análisis genético del comportamiento los sujetos con un factor de ascendencia genética (GAF) europea de 1. Estos 440 individuos de ascendencia europea (edad promedio de 11,5 [desviación estándar = 4,8] años, 53,0 % de sexo masculino) se analizaron utilizando el rendimiento cuantitativo en las puntuaciones de Reconocimiento de Lectura Oral y Vocabulario con Imágenes con PLINK v1.07, con la edad incluida como covariable (Purcell *et al.*, 2007). Para corregir las pruebas múltiples (20 pruebas totales = 10 SNP x 2 mediciones de lenguaje), establecimos umbrales estadísticos utilizando la tasa de descubrimiento falso con $\alpha = 0,05$ (Benjamini y Hochberg, 1995).

Análisis de imágenes en PING.

Las técnicas de obtención de imágenes, la adquisición de datos y los análisis de PING se tratan en profundidad en otro lugar y brevemente a continuación (Brown *et al.*, 2012; Fjell *et al.*, 2012; Walhovd *et al.*, 2012). En los diez sitios y 12 escáneres, se implementó un protocolo estandarizado de IRM estructural de alta resolución de modalidad múltiple, que implicó volúmenes 3D ponderados en T1 y T2, y un conjunto de exploraciones ponderadas por difusión. En la Universidad de California en San Diego, los datos se obtuvieron en un escáner GE 3T SignaHDx y un escáner 3T Discovery 750x (GE Healthcare) utilizando bobinas de cabeza en serie de ocho canales. El protocolo incluía un

localizador de tres planos convencional, una ecoimagen sagital en 3D con recuperación de inversión degradada ponderada en T1 con volumen optimizado para el máximo contraste de materia gris/blanca (tiempo de ecoimagen = 3,5 ms, tiempo de repetición = 8,1 ms, tiempo de inversión = 640 ms, ángulo de giro = 8°, ancho de banda del receptor = ± 31,25 kHz, FOV = 24 cm, frecuencia = 256, fase = 192, grosor del corte = 1,2 mm), y dos exploraciones de imágenes de tensor de difusión en 2D (DTI) axiales (30 direcciones valor b = 1.000, TE = 83 ms, TR = 13.600 ms, frecuencia = 96, fase = 96, espesor de corte = 2,5 mm). Se instalaron protocolos de adquisición con parámetros de secuencia de pulsos idénticos o casi idénticos a los protocolos utilizados en la Universidad de California en San Diego en escáneres en los otros nueve sitios. Se adquirieron datos en todos los escáneres para estimar las tasas de relajación, y medir y corregir el gradiente de deformación no lineal de la bobina específica del escáner. Los archivos de imagen en formato DICOM fueron procesados con un flujo de procesamiento automatizado escrito en MATLAB (Natick, MA) y C++ por el Laboratorio de Imágenes Multimodales de UCSD. Las imágenes estructurales ponderadas en T1 se corrigieron por distorsiones causadas por gradientes de no linealidad, se corrigieron, se promediaron y se volvieron a muestrear rígidamente en alineación con un atlas cerebral. El procesamiento y el análisis de las imágenes se realizaron utilizando un conjunto totalmente automatizado de herramientas disponibles en el paquete de software FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>), así como un método basado en atlas para delinear y etiquetar los tractos de fibra de WM (Fischl, 2012).

Imágenes del tensor de difusión.

En las imágenes ponderadas por difusión, se corrigió la distorsión de la corriente de Foucault utilizando un método de descenso de gradiente conjugado iterativo e inverso de mínimos cuadrados para resolver los 12 parámetros de escalado y traducción que describen las distorsiones de la corriente de Foucault en toda la exploración por IRM de difusión, teniendo en cuenta explícitamente las orientaciones y amplitudes del gradiente de difusión (Zhuang *et al.*, 2006). El movimiento de la cabeza se corrigió registrando cada imagen ponderada por difusión a una imagen correspondiente sintetizada desde un ajuste tensorial a los datos (Hagler *et al.*, 2009). Para los datos de la IRM de difusión, se corrigieron las distorsiones espaciales y de intensidad causadas por la falta de homogeneidad del campo magnético B0 mediante el método del gradiente de inversión (Holland *et al.*, 2010). Las distorsiones causadas por las no linealidades de gradiente se corrigieron aplicando una transformación no lineal, predefinida, específica del escáner (Jovicich *et al.*, 2006). Las imágenes ponderadas por difusión se registraron automáticamente en imágenes estructurales ponderadas en T1 utilizando información mutua (Wells *et al.*, 1996) y se volvieron a muestrear rígidamente en una orientación estándar con respecto a las imágenes ponderadas en T1 con vóxeles isotrópicos de 2 mm. Se utilizó interpolación cúbica para todos los pasos de remuestreo. Se utilizaron métodos DTI convencionales para calcular las medidas de difusión (Basser *et al.*, 1994; Pierpaoli *et al.*, 1996). La duración del escaneo para la secuencia DTI fue de 4:24 minutos. Los tractos de fibra de materia blanca se marcaron utilizando un método de segmentación basado en atlas probabilístico (Hagler *et al.*, 2009). Los vóxeles que contienen principalmente materia gris o líquido cefalorraquídeo, identificados mediante la segmentación cerebral automática de FreeSurfer, fueron excluidos del análisis (Fischl *et al.*, 2002). Los volúmenes de tractos de fibra se calcularon como el número de vóxeles con probabilidad mayor que 0,08, el valor que proporcionó una correspondencia óptima en volumen entre las regiones de interés derivadas del atlas y los tractos de fibra trazados manualmente.

Análisis estadísticos.

Los análisis de imagenología-genética se realizaron en individuos de ascendencia genética europea. El escáner, la edad, la lateralidad manual, el estatus socioeconómico y el sexo se incluyeron como covariables en todos los análisis (Akshoomoff *et al.*, 2013; Brown *et al.*, 2012; Fjell *et al.*, 2012; Walhovd *et al.*, 2012). 332 sujetos de ascendencia genética europea habían realizado mediciones de imágenes que pasaron el control de calidad PING. Los volúmenes de los tractos de fibra en 16 zonas de interés se probaron mediante análisis de regresión múltiple en R utilizando el portal de datos de PING (<https://mmil-dataportal.ucsd.edu>).

Resultados

SNP y asociaciones basadas en genes

En la Tabla 2, se presentan las diez asociaciones de GWAS más fuertes de DL y TL concomitantes en ALSPAC. Las asociaciones más fuertes se observaron con *ZNF385D* (OR = 1,81, $p = 5,45 \times 10^{-7}$) y *COL4A2* (OR = 1,71, $p = 7,59 \times 10^{-7}$) (Tabla 2). A continuación, se examinaron individualmente DL y TL, sin incluir casos concomitantes, para determinar si uno de los trastornos estaba impulsando estas asociaciones. Las diez asociaciones más fuertes para casos de DL y casos de TL se presentan individualmente en la Tabla 3 y la Tabla 4, respectivamente. Las asociaciones más fuertes con TL fueron con marcadores en *NDST4* (OR = 1,83, $p = 1,40 \times 10^{-7}$) (Tabla 3). Los marcadores en el cromosoma 10 (OR = 1,43, $p = 5,16 \times 10^{-6}$), el cromosoma 8 (OR = 1,70, $p = 5,85 \times 10^{-6}$) y el gen *OPA3* (OR = 1,53, $p = 6,92 \times 10^{-6}$) tenían las asociaciones más fuertes con DL (Tabla 4). Los marcadores con $p < 0,01$ dentro de los genes previamente implicados en DL y/o TL se presentan en la Tabla Suplementaria 1 para cada fenotipo. Las asociaciones más fuertes con estos marcadores se observaron para *KIAA0319* con DL y TL concomitantes (rs16889556, $p = 0,0005177$), *FOXP2* con DL y TL concomitantes (rs1530680, 0,0001702), *CNTNAP2* con TL (rs6951437, $p = 0,0000462$) y *DCD2* con TL (rs793834, 0,0002679) (Tabla Suplementaria 1a-1c). Se realizaron análisis genéticos en cada fenotipo (DL y TL concomitantes, DL individualmente y TL individualmente), y las diez asociaciones genéticas más sólidas se presentan en la Tabla Suplementaria 2. Ninguna de las

asociaciones genéticas sobrevivió a la corrección para pruebas múltiples; sin embargo, las asociaciones más fuertes fueron vistas con: (1) *OR5H2*, *OR5H6* y *RRAGA* con DL y TL concomitantes, (2) *NEK2*, *DLEC1* y *NARS* con TL, y (3) *MAP4*, *OR2L8* y *CRYBA4* con DL. Los marcadores con los valores p más fuertes en los análisis de descubrimiento en *ZNF385D*, *COL4A2* y *NDST4* fueron llevados hacia el análisis de replicación en PING. Observamos la replicación de dos marcadores dentro de *ZNF385D* y el rendimiento en la Prueba de Vocabulario con Imágenes ($p = 0,00245$ y $0,004173$) (Tabla 5). Sin embargo, los marcadores no se replicaron con la Prueba de Reconocimiento de Lectura Oral ($p > 0,05$).

Imagenología genética de *ZNF385D*

Para hacer un seguimiento de las asociaciones replicadas de *ZNF385D*, se examinaron los efectos de estas variantes en los volúmenes de tractos de fibra previamente implicados en el lenguaje escrito y verbal. Antes de hacerlo, el volumen de los tractos de fibra se determinó por primera vez como un factor predictivo del rendimiento en el Reconocimiento de Lectura Oral y las Pruebas de Vocabulario con Imágenes (datos no mostrados). En sujetos con antecedentes genéticos solo europeos, los genotipos de *ZNF385D* fueron predictores del volumen total de los tractos de fibra y también de los volúmenes de los tractos de fibra en los hemisferios derecho e izquierdo (Tabla 6). Los SNP de *ZNF385D* también fueron predictores bilateralmente dentro del fascículo longitudinal inferior (ILF), el fascículo frontooccipital inferior (IFO) y el fascículo temporal longitudinal superior (tSLF) en este subconjunto (Tabla 6). Para discernir si estas asociaciones entre *ZNF385D* y los volúmenes de los tractos de fibra reflejaban diferencias globales en el volumen cerebral según el genotipo, se examinó la relación de *ZNF385D* con la segmentación cerebral total y los volúmenes corticales totales. Se encontraron asociaciones para ambas medidas con rs1679255 ($p = 0,00072$ y $0,00027$, respectivamente) y rs12636438 ($p = 0,000259$ y $0,000069$, respectivamente). Los efectos parecían ser de naturaleza aditiva, con individuos heterocigotos que tienen fenotipos intermedios en relación con aquellos homocigotos para el alelo mayor y con aquellos homocigotos para el alelo menor. De hecho, cuando estas medidas del volumen cerebral total se insertaron en el modelo como covariable, ya no existían asociaciones de *ZNF385D* con los volúmenes de los tractos de fibra por DTI.

Como se describe en el presente documento, se identificaron genes que contribuyen a la aparición conjunta de DL y TL. En los análisis de descubrimiento, se encontraron asociaciones de *ZNF385D* y *COL4A2* en casos concomitantes, y de *NDST4* con TL. Después, se observaron asociaciones de *ZNF385D* con el rendimiento en una medición del vocabulario, pero no en una medición de la lectura oral, en PING. La asociación con el rendimiento en una medición del vocabulario, aunque no recapitula exactamente el fenotipo de comorbilidad, proporciona evidencia adicional de la contribución de *ZNF385D* al lenguaje. Para obtener una comprensión funcional, se interrogaron los efectos de los marcadores de *ZNF385D* replicados en los volúmenes de tractos de fibra relacionados con el lenguaje. Los marcadores de *ZNF385D* se asocian bilateralmente con los volúmenes generales de los tractos de fibra, así como con el volumen cerebral general.

Los estudios han demostrado que DL y TL comparten contribuyentes genéticos (Trzaskowski *et al.* 2013). Sin embargo, los genes específicos que contribuyen tanto a DL como a TL solo han comenzado a ser examinados recientemente. Estos estudios solo han usado un enfoque de gen candidato para examinar esta etiología genética compartida. Tal enfoque ha tenido éxito al mostrar la contribución compartida de *DCDC2*, *KIAA0319*, *FOXP2*, *CNTNAP2*, entre otros, a DL y TL (Eicher y Gruen, 2013; Graham y Fisher, 2013; Newbury *et al.*, 2009; Newbury *et al.*, 2010; Pintel *et al.*, 2012; Rice *et al.*, 2009; Scerri *et al.*, 2011). De hecho, los marcadores dentro de *KIAA0319*, *FOXP2* y *CNTNAP2* (junto con *BC0307918*) mostraron una asociación nominal con DL y TL concomitantes en los análisis ($p < 0,01$) descritos en el presente documento. Los genes de riesgo de DL/TL también mostraron una tendencia a asociarse con TL individualmente (*DCDC2*, *KIAA0319* y *CNTNAP2*) y con DL individualmente (*CNTNAP2* y *CMIP*) ($p < 0,01$). Es probable que la falta de replicación para otros genes de riesgo de DL/TL, y las diferencias específicamente entre este estudio y los de Scerri *et al.* (2011) y Luciano *et al.* (2013) sean resultado de diferentes definiciones de casos y números, ya que se diseñaron clasificaciones de casos instantáneas para capturar el mayor rango posible de sujetos con problemas de lectura y lenguaje en lugar de usar medidas neurocognitivas altamente específicas.

Una omisión evidente en las investigaciones genéticas de DL y TL es la falta de métodos libres de hipótesis. Estos métodos permiten el descubrimiento de nuevos genes porque no dependen de candidatos preseleccionados. Aquí, los análisis de GWAS indican que *ZNF385D* contribuye a la comorbilidad de DL y TL. Este estudio no es el primero en realizar un GWAS en rasgos relacionados con la lectura y el lenguaje. Luciano *et al.* (2013) informaron recientemente de un GWAS de mediciones cuantitativas del lenguaje verbal y escrito en dos cohortes basadas en la población, incluida ALSPAC. Encontraron pruebas sólidas de que *ABCC13*, *BC0307918*, *DAZAP1*, entre otros, afectan el rendimiento en estas mediciones, aunque los análisis instantáneos no proporcionaron pruebas sólidas para ello. Las estrategias analíticas diferían de dos maneras: (1) el uso de medidas dicotómicas en lugar de cuantitativas para condicionar las asociaciones genéticas y (2) el análisis de la lectura y el lenguaje juntos en lugar de hacerlo de manera individual. Los estudios de asociación anteriores de DL y TL han mostrado diferencias en los resultados dependiendo de si los datos genéticos estaban condicionados por fenotipos dicotómicos o cuantitativos. Por ejemplo, *KIAA0319* tiende a asociarse más fácilmente con mediciones cuantitativas, mientras que *DCDC2* se asocia más a menudo con variables dicotomizadas (Paracchini *et al.*, 2008; Powers *et al.*, 2013; Scerri *et al.*, 2011). El presente estudio, que examina la comorbilidad, y el de Luciano *et al.*, que examinó el rendimiento en tareas de lectura y lenguaje de forma individual, condicionó asociaciones genéticas en diferentes rasgos, lo que puede llevar a

diferentes asociaciones estadísticas. Ambas estrategias analíticas son válidas y han recogido información distinta, pero relacionada, sobre los fundamentos genéticos del lenguaje verbal y escrito. Demuestran la importancia del examen creativo y cuidadoso de los fenotipos al examinar los rasgos neurocognitivos y otros rasgos complejos.

5 Tras el análisis primario de DL y TL concomitantes, se examinaron DL y TL individualmente para determinar si un solo trastorno estaba impulsando las señales de asociación. *ZNF385D* no se asoció ni con DL ni con TL individualmente, lo que indica que *ZNF385D* contribuye a los procesos relacionados tanto con DL como con TL, en vez de contribuir a solo uno de estos trastornos. Dentro de la cohorte de PING, se observaron asociaciones de marcadores de *ZNF385D* con el rendimiento en la Prueba de Vocabulario con Imágenes y no en la Prueba de Reconocimiento de Lectura Oral. Las mediciones del vocabulario receptivo (por ejemplo, la Prueba de Vocabulario con Imágenes) están relacionadas con las tareas del lenguaje verbal y escrito (Scarborough 1990, Wise *et al.*, 2007), mientras que el rendimiento en las mediciones de la decodificación (por ejemplo, la Prueba de Reconocimiento de Lectura Oral) parece ser específico de la lectura. Por lo tanto, la Prueba de Vocabulario con Imágenes puede reflejar el fenotipo de DL y TL concomitantes utilizado para la asociación en ALSPAC mejor que la Prueba de Reconocimiento de Lectura Oral y explicar el patrón de asociación de *ZNF385D* en PING. Además de *ZNF385D*, se observaron asociaciones sugestivas de *COL4A2* con DL/TL concomitantes y de *NDST4* con TL. Ninguna de estas asociaciones se replicó con las mediciones en PING, pero los estudios que se realicen en el futuro deberían intentar replicar estas asociaciones, en particular debido a la participación conocida de *COL4A2* en lesiones pornecefálicas y de la sustancia blanca (Verbeek *et al.*, 2012, Yoneda *et al.*, 2011).

20 Los análisis genéticos no revelaron ninguna asociación que sobreviviera a la corrección para pruebas múltiples. No obstante, hubo asociaciones génicas intrigantes que deberían investigarse en estudios que se realicen en el futuro. Por ejemplo, con TL, hubo asociaciones sugestivas con varios genes en el cromosoma 19-*IL4I1*, *ATF5*, *NUP62* y *SIGLEC11*, que pueden corresponder al pico de enlace SLI2 (Monaco, 2007; SLI Consortium, 2002), Luciano *et al.* (2013) encontraron una acumulación similar de genes sugestivamente asociados a aproximadamente 5 Mb de los genes identificados en el presente documento. Además, *MAP4*, un gen de ensamblaje de microtúbulos, fue el gen que tuvo la asociación más fuerte con DL. Existe evidencia de que la función de los microtúbulos desempeña un papel clave en el desarrollo de la lectura, ya que se cree que la migración neuronal aberrante contribuye a la etiología de DL, y se cree que otros genes candidatos de DL interactúan con los microtúbulos (por ejemplo, *DCDC2* y *ACOT13*) (Cheng *et al.*, 2006). Estos hallazgos se pueden validar en una cohorte independiente, utilizando métodos descritos en el presente documento y métodos conocidos para concluir que están involucrados en DL y TL.

30 Las asociaciones más fuertes observadas en este estudio fueron con marcadores dentro de *ZNF385D*. *ZNF385D* se ha relacionado previamente a la esquizofrenia y al trastorno por déficit de atención con hiperactividad (TDAH) (Poelmans *et al.*, 2011; Xu *et al.*, 2013). Tanto la esquizofrenia como el TDAH son trastornos neuroconductuales que se cree que tienen alteraciones del núcleo en común con DL y TL, incluida la comprensión y el procesamiento semántico (Gilger *et al.*, 1992; Li *et al.*, 2009; Willcutt *et al.*, 2005). Además, la asociación observada de *ZNF385D*, como se describe en el presente documento, en el volumen cerebral global puede indicar que *ZNF385D* influye en varios rasgos neurocognitivos por medio de su efecto en todo el cerebro.

40 Se sabe poco acerca de la función de *ZNF385D*, aunque su dominio de dedos de zinc sugiere que es un regulador transcripcional. La importancia de la regulación transcripcional en el lenguaje verbal y escrito no es un concepto nuevo. El gen del lenguaje más ampliamente estudiado, *FOXP2*, es un potente factor de transcripción que se ha demostrado que regula otro gen del lenguaje, *CNTNAP2* (Vernes *et al.*, 2007; Vernes *et al.*, 2011). Además, en el locus *DYX2*, dos variantes de riesgo, *READ1* en *DCDC2* y el haplotipo de riesgo de *KIAA0319*, parecen tener la capacidad de regular la expresión de los genes (Couto *et al.*, 2010; Dennis *et al.*, 2009; Meng *et al.*, 2011) y, posiblemente, de interactuar (Ludwig *et al.* 2008; Ejemplo 1; Powers *et al.*, 2013). Las variantes de *ZNF385D* ahora se unen a esta lista de variantes transcripcionales putativas que influyen en las habilidades del lenguaje verbal y escrito. La caracterización de los genes objetivo de *ZNF385D* y de sus efectos transcripcionales en estos objetivos será un próximo paso importante. Además, la identificación de genes objetivo puede generar candidatos terapéuticos para el tratamiento y la cura de DL y TL. Para obtener más información sobre *ZNF385D*, se realizaron análisis genéticos imagenológicos de *ZNF385D* y de volúmenes de tractos de fibra de tractos relacionados con el lenguaje. *ZNF385D* parece modular los volúmenes de tractos de fibra y cerebral totales, que posteriormente pueden afectar la conectividad y la funcionalidad de las regiones cerebrales importantes en la integración eficiente y fluida del lenguaje escrito y verbal. Por lo tanto, la identificación de los genes objetivo y cómo la modulación de su expresión durante el desarrollo neuronal produce diferencias en los volúmenes de tractos de fibra y cerebral total será de suma importancia para diseccionar no solo el mecanismo de *ZNF385D*, sino también para el desarrollo de habilidades lingüísticas básicas en los niños.

55 Características de la población. En primer lugar, aunque el tamaño total de la muestra de ALSPAC sea formidable, el número de casos para cada definición es relativamente pequeño. Esto es esperable en una cohorte transversal de la población general, ya que la prevalencia de estos trastornos varía entre el 5 y el 17 % (Pennington y Bishop, 2009). No se esperaría que la cohorte de ALSPAC se enriquezca para casos de DL y/o TL. El tamaño pequeño de la muestra puede haber afectado la fuerza estadística y la capacidad de identificar genes de riesgo con un tamaño de efecto pequeño. En segundo lugar, las mediciones de la lectura y el lenguaje realizadas en los estudios de ALSPAC y PING no fueron idénticas. Los fenotipos en PING se trataron como un rasgo cuantitativo en lugar de tratarse como una variable dicotómica, como en ALSPAC. Por lo tanto, los intentos de replicar las asociaciones observadas en la

- cohorte de ALSPAC pueden haberse visto obstaculizados, ya que las mediciones de la lectura/el lenguaje en PING pueden haber capturado habilidades diferentes a las de ALSPAC. Sin embargo, las asociaciones observadas en PING indican que *ZNF385D* desempeña un papel importante y consistente en los procesos generales del lenguaje.
- 5 En tercer lugar, las mediciones de los volúmenes de tractos derivadas del atlas, así como los volúmenes derivados de tractos de fibra trazados manualmente, son probablemente subestimaciones del volumen de fibra real para la mayoría de los tractos. Sin embargo, los volúmenes de tractos de fibra se obtuvieron de manera consistente para todos los sujetos y, probablemente, reflejen diferencias interindividuales. No obstante, la fuerza y la replicación independiente de las asociaciones descritas en el presente documento y la relación con los fenotipos de imágenes cerebrales implican fuertemente a *ZNF385D* en los procesos del lenguaje central subyacentes a DL y TL.
- 10 En conclusión, *ZNF385D* fue identificado como un gen novedoso que contribuye tanto a DL como a TL, así como también al volumen de tracto de fibra y cerebral total. La implicación de otro factor de transcripción en los trastornos comunicativos subraya la importancia de la regulación transcripcional en el desarrollo neuronal de los dominios del lenguaje en el cerebro. Los estudios que se realicen en el futuro deben apuntar a caracterizar mejor la funcionalidad molecular de *ZNF385D* y replicar esta asociación, así como nuestras asociaciones no replicadas (*NDST4* y *COL4A2*) en DL, TL y otros trastornos relacionados.
- 15

Referencias

- Akshoomoff N., *et al.* (2013) *J Int Neuropsychol Soc*, bajo revisión.
- Alexander D.H., *et al.* (2009) *Genome Res* 19(9), 1655-64.
- Basser P.J., *et al.* (1994) *Biophys J* 66(1), 259-267.
- 20 Benjamini Y., *et al.* (1995) *J R Statst Soc B* 57(1): 289-300.
- Boyd A., *et al.* (2012) *Int J Epidemiol* 42(1), 111-27.
- Brown T.T., *et al.* (2012) *Curr Biol* 22(18), 1693-8.
- Catts H.W., *et al.* (2005) *J Speech Lang Hear Res* 48(6), 1378-96.
- Cheng Z., *et al.* (2006) *Biochem Biophys Res Commun* 350(4), 850-3.
- 25 Cope N., *et al.* (2012) *Neuroimage* 63(1), 148-56.
- Couto J.M., *et al.* (2010) *Am J Med Genet B Neuropsychiatr Genet* 153B(2), 447-62.
- Darki F., *et al.* (2012) *Biol Psychiatry* 72(8), 671-6.
- Dennis, M.Y., *et al.* (2009) *PLoS Genet* 5, e1000436.
- Eicher J.D. t Gruen J.R. (2013) *Mol Genet Metab*, doi: 10.1016/j.ymgme.2013.07.001.
- 30 Fischl B. (2012) FreeSurfer. *Neuroimage* 62(2), 774-81.
- Fischl B., *et al.* (2002) *Neuron* 33(3), 41-55.
- Fjell A.M., *et al.* (2012) *Proc Natl Acad Sci USA* 109(48), 19620-5.
- Gathercole S. y Baddeley A.D. (1990) *Journal of Memory and Language* 29, 336-360.
- Gathercole S.E. y Baddeley A.D. (1996) *The Psychological Corporation*, Londres.
- 35 Gilger J.W., *et al.* (1992) *JAm Acad Child Adolesc Psychiatry* 31(2), 343-8.
- Golding J., *et al.* (2001) I. Study methodology. *Paediatr Perinat Epidemiol* 15(1), 74-87.
- Graham S.A. y Fisher S.E. (2013) *Curr Opin Neurobiol* 23(1), 43-51.
- Hagler DJ, Jr., *et al.* (2009) *Hum Brain Mapp* 30(5): 1535-1547.
- Holland D., *et al.* (2010) *Neuroimage* 50 (1), 175-183.
- 40 Jovicich J., *et al.* (2006) *Neuroimage* 30(2): 436-443.
- Li X., *et al.* (2009) *Curr Opin Psychiatry* 22(2), 131-9.
- Liegeois F., *et al.* (2003) *Nat Neurosci* 6(11), 1230-7.

- Liu J.Z., *et al.* (2010) *Am J Hum Genet* 87(1), 139-45.
- Luciano M., *et al.* (2013) *Genes Brain Behav*, doi: 10.1111/gbb.12053.
- Ludwig K.U., *et al.* (2008) *J Neural Transm* 115(11), 1587-9.
- Meng H., *et al.* (2011) *Behav Genet* 41(1), 58-66.
- 5 Monaco A.P. (2007) *Ann Hum Genet* 71(Pt5), 660-73.
- Newbury D.F., *et al.* (2009) *Am J Hum Genet* 85(2), 264-72.
- Newbury D.F., *et al.* (2010) *Behav Genet* 41(1), 90-104.
- Neale M.D. (1997) *Neale Analysis of Reading Ability-Revised: Manual for Schools*, NFER-Nelson.
- Paracchini S., *et al.* (2008) *Am J Psychiatry* 165(12), 1576-84.
- 10 Peterson R.L. y Pennington, B.F. (2012) *Lancet* 379, 1997-2007.
- Pennington B.F. y Bishop D.V. (2009) *Annu Rev Psychol* 60, 283-306.
- Pennington B.F. (2006) *Cognition* 101(2), 385-413.
- Pierpaoli C., *et al.* (1996) *Radiology* 201(3), 637-648.
- Pinel P., *et al.* (2012) *J Neurosci* 32(3), 817-25.
- 15 Poelmans G., *et al.* (2011) *Am J Psychiatry* 168(4), 365-77.
- Powers N.R., *et al.* (2013) *Am J Hum Genet* 93(1), 19-28.
- Purcell S., *et al.* (2007) *Am J Hum Genet* 81(3), 559-575.
- Rice M.L., *et al.* (2009) *J Neurodev Disord* 1(4), 264-82.
- Rosner J. y Simon D.P. (1971) *Journal of Learning Disabilities* 4(384), 40-48.
- 20 Rust J., *et al.* (1993) *WORD: Wechsler Objective Reading Dimensions Manual*. Psychological Corporation, Sidcup, Reino Unido.
- Scarborough H.S. (1990) *Child Dev* 61(6), 1728-43.
- Scerri T.S., *et al.* (2012) *PLoS One* 7(11), e50312.
- Scerri T.S., *et al.* (2011) *Biol Psychiatry* 70(3), 237-45.
- 25 Scerri T.S. y Schulte-Korne G. (2010) *Eur Child Adolesc Psychiatry* 19(3), 179-97.
- Scott-Van Zeeland A.A., *et al.* (2010) *Sci Transl Med* 2(56), doi: 10.1126/scitranslmed.3001344.
- Shaywitz S.E. y Shaywitz B.A. (2008) *Dev Psychopathol* 20(4), 1329-49.
- SLI Consortium. (2002) *Am J Hum Genet* 70(2), 384-98.
- Tan G.C., *et al.* (2010) *Neuroimage* 53(3), 1030-42.
- 30 Trzaskowski M., *et al.* (2013) *Behav Genet* 43(4), 267-73.
- Vandermosten M., *et al.* (2012) *Brain* 135(Pt 3), 935-48.
- Verbeek E., *et al.* (2012) *Eur J Hum Genet* 20(8), 844-51.
- Vernes S.C., *et al.* (2011) *PLoS Genet* 7(7), e1002145.
- Vernes S.C., *et al.* (2007) *Am J Hum Genet* 81(6), 1232-50.
- 35 Walhovd K.B., *et al.* (2012) *Proc Natl Acad Sci USA* 109(49), 20089-94.
- Wechsler D. (1996) *Wechsler objective language dimensions (WOLD)*. The Psychological Corporation, Londres.

Wechsler D., *et al.* (1992) *WISC-IIIUK: Wechsler Intelligence Scale for Children*. Psychological Corporation, Sidcup, Reino Unido.

Weintraub S., *et al.* (2013) *Neurology* 90(11 Supl. 3), S54-64.

Wells W.M. 3.^o, *et al.* (1996) *Med Image Anal* 1(1), 35-51.

5 Wilcke A., *et al.* (2011) *Eur J Hum Genet* 20(2), 224-9.

Willcutt E.G., *et al.* (2005) *Dev Neuropsychol* 27(1), 35-78.

Wise J.C., *et al.* (2007) *J Speech Lang Hear Res* 50(4), 1093-9.

Xu C., *et al.* (2013) *PLoS One* 8(1), e51674.

Yoneda Y., *et al.* (2012) *Am J Hum Genet* 90(1), 86-90.

10 Zhuang J., *et al.* (2006)

Tablas del Ejemplo 4

Tabla 1: Mediciones de la lectura y el lenguaje utilizadas para definir los casos de discapacidad lectora (DL) y trastorno del lenguaje (TL)

Discapacidad lectora (DL) (n = 527)*	Trastorno del lenguaje (TL) (n = 337)**
Eliminación de fonemas Edad 7 años	Eliminación de fonemas Edad 7 años
Lectura de palabras sueltas Edad 7 años	Comprensión verbal Edad 8 años
Lectura de palabras sueltas Edad 9 años	Repetición de seudopalabras Edad 8 Años
Lectura de seudopalabras Edad 9 años	
Comprensión lectora Edad 9 años	

*Los casos de DL tuvieron una puntuación z menor o igual a -1 en al menos 3 de las 5 mediciones de la lectura.
 **Los casos de TL tuvieron una puntuación z menor o igual a -1 en al menos 2 de las 3 mediciones del lenguaje.

15 Tabla 2: Asociaciones con casos de DL y TL concomitantes en ALSPAC (n = 174)

Marcador	Cro	Par de base	Alelo menor	FAM Af	FAM No Af	Gen	Ratio de probabilidades	Valor p
rs12636438	3	22038281	G	0,3017	0,1927	ZNF385D	1,811	5,45 x 10 ⁻⁷
rs1679255	3	22022938	C	0,3006	0,1923	ZNF385D	1,805	6,87 x 10 ⁻⁷
rs9521789	13	109917621	C	0,5201	0,3879	COL4A2	1,71	7,59 x 10 ⁻⁷
rs1983931	13	109916103	G	0,5201	0,3896	COL4A2	1,698	1,06 x 10 ⁻⁶
rs9814232	3	21948179	A	0,2931	0,1886	ZNF385D	1,784	1,30 x 10 ⁻⁶
rs7995158	13	109909718	A	0,5201	0,3911		1,687	1,44 x 10 ⁻⁶
rs6573225	14	58354640	C	0,1965	0,1122		1,935	1,56 x 10 ⁻⁶
rs4082518	10	17103032	T	0,3103	0,2049	CUBN	1,746	2,17 x 10 ⁻⁶

ES 2 718 343 T3

Marcador	Cro	Par de base	Alelo menor	FAM Af	FAM No Af	Gen	Ratio de probabilidades	Valor p
rs442555	14	58365937	C	0,1983	0,1149		1,905	2,38 x 10 ⁻⁶
rs259521	3	21942154	T	0,2902	0,1885	<i>ZNF385D</i>	1,761	2,42 x 10 ⁻⁶

Cro, cromosoma; FAM Af, frecuencia de alelos menores en sujetos afectados; FAM No Af, frecuencia de alelos menores en sujetos no afectados.

Tabla 3: Asociaciones con casos de TL en ALSPAC, excluyendo casos de DL concomitantes (n = 163)

Marcador	Cro	Par de base	Alelo menor	FAM Af	FAM No Af	Gen	Ratio de probabilidades	Valor p
rs482700	4	116286939	G	0,3896	0,2588	<i>NDST4</i>	1,827	1,40 x 10 ⁻⁷
rs7695228	4	116309516	T	0,3920	0,2636	<i>NDST4</i>	1,801	2,94 x 10 ⁻⁷
rs1940309	4	116306410	T	0,3865	0,2606	<i>NDST4</i>	1,788	4,14 x 10 ⁻⁷
rs505277	4	116248257	T	0,3773	0,2528	<i>NDST4</i>	1,791	4,35 x 10 ⁻⁷
rs476739	4	116248997	A	0,3773	0,2529	<i>NDST4</i>	1,79	4,41 x 10 ⁻⁷
rs867036	4	116381578	C	0,3957	0,2696	<i>NDST4</i>	1,774	5,31 x 10 ⁻⁷
rs867035	4	116381423	C	0,3957	0,2697	<i>NDST4</i>	1,773	5,45 x 10 ⁻⁷
rs2071674	4	2366882	T	0,0920	0,0389	<i>ZFYVE28</i>	2,503	1,90 x 10 ⁻⁶
rs7694946	4	116413588	C	0,3620	0,2526	<i>NDST4</i>	1,678	8,95 x 10 ⁻⁶
rs4823324	22	44616787	C	0,2914	0,4143	<i>ATXN10</i>	0,581	9,30 x 10 ⁻⁶

Cro, cromosoma; FAM Af, frecuencia de alelos menores en sujetos afectados; FAM No Af, frecuencia de alelos menores en sujetos no afectados.

Tabla 4: Asociaciones con casos de DL en ALSPAC, excluyendo casos de TL concomitantes (n = 353)

Marcador	Cro	Par de base	Alelo menor	FAM Af	FAM No Af	Gen	Ratio de probabilidades	Valor p
rs180950	10	115697957	G	0,456	0,369		1,431	5,16 x 10 ⁻⁶
rs2590673	8	126037337	G	0,133	0,083		1,697	5,85 x 10 ⁻⁶
rs892100	19	50772522	C	0,228	0,162	<i>OPA3</i>	1,526	6,92 x 10 ⁻⁶
rs1792745	18	51955991	T	0,187	0,129		1,558	1,22 x 10 ⁻⁵
rs12546767	8	126151747	C	0,152	0,099	<i>KIAA0196</i>	1,618	1,32 x 10 ⁻⁵

Marcador	Cro	Par de base	Alelo menor	FAM Af	FAM No Af	Gen	Ratio de probabilidades	Valor p
rs12634033	3	146524529	C	0,135	0,087		1,646	1,80 x 10 ⁻⁵
rs892270	12	105002956	G	0,534	0,451	<i>NUAK1</i>	1,395	2,16 x 10 ⁻⁵
rs10887149	10	124156994	A	0,278	0,357	<i>PLEKHA1</i>	0,069	2,25 x 10 ⁻⁵
rs10041417	5	33218502	T	0,226	0,164		1,489	2,58 x 10 ⁻⁵
rs6792971	3	68468217	C	0,111	0,068	<i>FAM19A1</i>	1,703	2,59 x 10 ⁻⁵

Cro, cromosoma; FAM Af, frecuencia de alelos menores en sujetos afectados; FAM No Af, frecuencia de alelos menores en sujetos no afectados.

Tabla 5: Replicación de asociaciones en PING (n = 440)

Marcador	Alelo menor	FAM	Gen	Prueba de lectura oral		Prueba de vocabulario con imágenes	
				Beta	Valor p	Beta	Valor p
rs12636438	G	0,161	<i>ZNF385D</i>	-0,1867	0,9452	-2,88	0,004173*
rs1679255	G	0,292	<i>ZNF385D</i>	-1,84	0,5016	-3,048	0,002445**
rs9521789	G	0,4370	<i>COL4A2</i>	-0,3411	0,7332	0,8647	0,3877
rs476739	A	0,265	<i>NDST4</i>	0,5406	0,5891	0,5159	0,6062
rs505277	A	0,280	<i>NDST4</i>	0,5406	0,5891	-0,3452	0,7301
rs482700	G	0,278	<i>NDST4</i>	0,5498	0,5828	-0,05341	0,9574
rs7695228	A	0,295	<i>NDST4</i>	0,6258	0,5318	0,09991	0,9205
rs867036	G	0,378	<i>NDST4</i>	0,2605	0,7946	-0,1414	0,8876
rs867035	G	0,377	<i>NDST4</i>	0,2961	0,7673	-0,1565	0,8757
rs1940309	A	0,281	<i>NDST4</i>	0,6049	0,5456	0,1296	0,8969

*Valor p menor que el umbral estadístico ajustado por FDR (umbral ajustado por FDR = 0,05 x (2/19) = 0,00526

**Valor p menor que el umbral estadístico ajustado por FDR (umbral ajustado por FDR = 0,05 x (1/20) = 0,00250

FAM, frecuencia de alelos menores en una muestra completa de PING

5 Tabla 6: Asociaciones de *ZNF385D* con volúmenes de tractos de fibra de DTI en sujetos con un 100 % de ascendencia genética europea (n = 332)

Tracto de fibra	rs1679255	rs12636438
-----------------	-----------	------------

ES 2 718 343 T3

	Pendiente	Valor p	Pendiente	Valor p
Todos	-3329,9	0,044*	-3717,9	0,023*
Todos derecha	-1731,4	0,039*	-1965	0,017 *
Todos izquierda	-1616,3	0,055	-1775,6	0,033*
FLI derecho	-251,3	0,011*	-234,4	0,016*
FLI izquierdo	-256,9	0,0088**	-254,6	0,009**
FOI derecho	-200,8	0,032*	-190	0,041*
FOI izquierdo	-221	0,012*	-226,3	0,009**
FLS derecho	-168,1	0,06	-206	0,02*
FLS izquierdo	-199,5	0,022*	-212,9	0,013*
FLSt derecho	-170,8	0,011*	-180,7	0,0068**
FLSt izquierdo	-163,1	0,023*	-169,9	0,016*
FLSp derecho	-153,1	0,079	-182,4	0,034*
FLSp izquierdo	-112,2	0,18	-125,3	0,131
CFIE	-148,8	0,052	-165,6	0,029*
CFIE izquierda	-34,54	0,66	-54,3	0,48
CC	-977,1	0,15	-1181,6	0,081
*p ≤0,05				
**p ≤0,01				
Abreviaturas: Todos (todos los tractos de fibra), FLI (fascículo longitudinal inferior), FOI (fascículo frontooccipital inferior), FLS (fascículo longitudinal superior), FLSt (fascículo longitudinal superior temporal), FLSp (fascículo longitudinal superior parietal), CFIE (corteza frontal inferior estriatal), CC (cuerpo caloso)				

Tablas suplementarias

Tabla suplementaria 1: Asociaciones de marcadores dentro de genes previamente implicados en DL y/o TL con (a) TL y DL concomitantes, (b) TL individualmente y (c) DL individualmente.

5 a) DL y TL concomitantes

Marcador	Gen	Cro	Par de base	Valor p
rs16889556	<i>KIAA0319</i>	6	24749584	0,0005177
rs1047782	<i>TDP2</i>	6	24758710	0,006515

ES 2 718 343 T3

Marcador	Gen	Cro	Par de base	Valor p
rs1530680	<i>FOXP2</i>	7	114194632	0,0001702
rs12667130	<i>FOXP2</i>	7	114213035	0,003033
rs6965855	<i>CNTNAP2</i>	7	145348483	0,006804
rs985080	<i>CNTNAP2</i>	7	145359118	0,006157
rs4726782	<i>CNTNAP2</i>	7	145425012	0,005341
rs1718101	<i>CNTNAP2</i>	7	145753721	0,0008707
rs10487689	<i>CNTNAP2</i>	7	146835482	0,008787
rs1918296	<i>CNTNAP2</i>	7	147655135	0,00616
rs737533	<i>BC0307918</i>	10	3353137	0,001008

b) TL

Marcador	Gen	Cro	Par de base	Valor p
rs793845	<i>DCDC2</i>	6	24296970	0,005511
rs2799373	<i>DCDC2</i>	6	24303738	0,0009664
rs793862	<i>DCDC2</i>	6	24315179	0,002443
rs793834	<i>DCDC2</i>	6	24342912	0,0002679
rs2792682	<i>DCDC2</i>	6	24380363	0,006634
rs807704	<i>DCDC2</i>	6	24408825	0,001988
rs707864	<i>DCDC2</i>	6	24413827	0,001266
rs12193738	<i>KIAA0319</i>	6	24676372	0,00974
rs2817198	<i>KIAA0319</i>		24683073	0,00559
rs10456309	<i>KIAA0319</i>	6	24697541	0,002258
rs985080	<i>CNTNAP2</i>	7	145359118	0,006735
rs1554690	<i>CNTNAP2</i>	7	145377266	0,006486
rs2533096	<i>CNTNAP2</i>	7	146037312	0,004782
rs6951437	<i>CNTNAP2</i>	7	146037340	0,0000462
rs344470	<i>CNTNAP2</i>	7	146044430	0,001697

ES 2 718 343 T3

Marcador	Gen	Cro	Par de base	Valor p
rs344468	<i>CNTNAP2</i>	7	146050259	0,003965

c) DL

Marcador	Gen	Cro	Par de base	Valor p
rs4725745	<i>CNTNAP2</i>	7	147032172	0,002407
rs12444778	<i>CMIP</i>	16	80330728	0,003148
rs1444186	<i>CMIP</i>	16	80330745	0,00482

Tabla suplementaria 2: Análisis genéticos de DL y TL concomitantes, TL individualmente y DL individualmente. Se muestran las diez principales asociaciones genéticas para cada uno.

Gen	Cr	Posición de inicio	Posición de parada	N.º SNP en gen	Valor p
DL y TL					
<i>OR5H2</i>	3	99484421	99485366	16	0,000072
<i>OR5H6</i>	3	99465818	99466796	19	0,000127
<i>RRAGA</i>	9	19039371	19041021	30	0,000276
<i>OR6B3</i>	2	240633166	240634162	36	0,000294
<i>UMOD</i>	16	20251873	20271538	29	0,000307
<i>A26C1A</i>	2	131692393	131738886	1	0,000389
<i>FAM29A</i>	9	19043140	19092902	44	0,000406
<i>CHRNA1</i>	2	175320568	175337446	23	0,000420
<i>IFIT5</i>	10	91164418	91170733	27	0,000475
<i>LOC643905</i>	2	240629902	240631072	39	0,000562
TL					
<i>NEK2</i>	1	209902744	209915590	28	0,000117
<i>DLEC1</i>	3	38055699	38139232	20	0,000171
<i>NARS</i>	18	53418891	53440175	36	0,000203
<i>IL4I1</i>	19	55084722	55124574	22	0,000305
<i>PKD2</i>	4	89147843	89217953	34	0,000313

ES 2 718 343 T3

Gen	Cr	Posición de inicio	Posición de parada	N.º SNP en gen	Valor p
DL y TL					
<i>ATF5</i>	19	55123785	55129004	18	0,000344
<i>NUP62</i>	19	55101893	55124598	19	0,000402
<i>SIGLEC11</i>	19	55144061	551556241	49	0,000578
<i>ACAN</i>	15	87147677	87219589	43	0,000633
<i>PGD</i>	1	10381671	10402788	12	0,000668
DL					
<i>MAP4</i>	3	47867188	48105715	18	0,000085
<i>OR2L8</i>	1	246178782	246179721	19	0,000139
<i>CRYBA4</i>	22	25347927	25356636	40	0,000219
<i>OR2T8</i>	1	246150942	246151881	24	0,000225
<i>KIAA1622</i>	14	93710401	93815825	42	0,000255
<i>OR2AK2</i>	1	246195256	246196264	15	0,000315
<i>DHX30</i>	3	47819654	47866687	11	0,000316
<i>GEMIN6</i>	2	38858830	38862610	8	0,000351
<i>C20orf10</i>	20	43435933	43440371	23	0,000450
<i>PPIF</i>	10	80777225	38862610	22	0,000493

Listado de secuencias

- 5 <110> Universidad de Yale
Gruen, Jeffrey R
- <120> EVALUACIÓN DEL RIESGO DE DISCAPACIDAD LECTORA Y LINGÜÍSTICA
- <130> Y0087.70061WO00
- <140> TBD
<141> 2013-11-07
- 10 <150> US 61/723,774
<151> 2012-11-07
- <160> 68
- <170> PatentIn versión 3.5

<210> 1
 <211> 38
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> Oligonucleótido sintético

<400> 1
 tgtaaacga cggccagtg ttgaatcca gaccaca 38

10

<210> 2
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

15

<400> 2
 atcccgatga aatgaaagg 20

20

<210> 3
 <211> 18
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 3
 tgtaaacga cggccagt 18

25

<210> 4
 <211> 20
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> Oligonucleótido sintético

<400> 4
 agcctgccta ccacagagaa 20

35

<210> 5
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

40

<400> 5
 ggaacaacct cacagaaatg g 21

<210> 6
 <211> 21
 <212> ADN
 <213> Secuencia artificial

ES 2 718 343 T3

<220>
 <223> Oligonucleótido sintético

<400> 6
 tgaaaccccg tctctactga a 21

5

<210> 7
 <211> 47
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> Oligonucleótido sintético

<400> 7
 ttgagaggaa ggaaggaag gatccctgag aggaaggaaa ggaagga 47

15

<210> 8
 <211> 47
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> Oligonucleótido sintético

<400> 8
 aatccttctc ttcttctc tcagggatcc ttcttctc tcctctc 47

25

<210> 9
 <211> 47
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 9
 ttgagagaga gagagagaga gatccctgag agagagagag agagaga 47

30

<210> 10
 <211> 47
 <212> ADN
 <213> Secuencia artificial

35

<220>
 <223> Oligonucleótido sintético

<400> 10
 aatctctc tcctctctc tcagggatct ctctctctc ctctctc 47

40

<210> 11
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 11
 tcatgcaaag ttccaaaacc 20

<210> 12
 <211> 19
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> Oligonucleótido sintético

<400> 12
 gatttcctcc ctcccttc 19

10

<210> 13
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

15

<400> 13
 gccctaggca ccagggtgtg a 21

20

<210> 14
 <211> 20
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 14
 acagggtgct cctcaggggc 20

25

<210> 15
 <211> 26
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <223> Oligonucleótido sintético

<400> 15
 gagaggaagg aaagagagga aggaaa 26

35

<210> 16
 <211> 13
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

40

<400> 16
 gagaggaagg aaa 13

<210> 17
 <211> 26
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 17
 gagaggaagg aaagagagga agaaaa 26

5

<210> 18
 <211> 25
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> Oligonucleótido sintético

<400> 18
 gagaggaagg aaagagagga aggaa 25

15

<210> 19
 <211> 28
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> Oligonucleótido sintético

<400> 19
 ggaaggaagg aaggaaggaa ggaaggaa 28

25

<210> 20
 <211> 36
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 20
 ggaaggaagg aaggaaggaa ggaaggaagg aaggaa 36

30

<210> 21
 <211> 24
 <212> ADN
 <213> Secuencia artificial

35

<220>
 <223> Oligonucleótido sintético

<400> 21
 ggaaggaagg aaggaaggaa ggaa 24

40

<210> 22
 <211> 32
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Oligonucleótido sintético

<400> 22
 ggaaggaagg aaggaaggaa ggaaggaagg aa 32

ES 2 718 343 T3

	<210> 23		
	<211> 16		
	<212> ADN		
	<213> Secuencia artificial		
5	<220>		
	<223> Oligonucleótido sintético		
	<400> 23		
	ggaaggaagg aaggaa	16	
10	<210> 24		
	<211> 20		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Oligonucleótido sintético		
15	<400> 24		
	ggaaggaagg aaggaaggaa	20	
	<210> 25		
	<211> 40		
	<212> ADN		
20	<213> Secuencia artificial		
	<220>		
	<223> Oligonucleótido sintético		
	<400> 25		
	ggaaggaagg aaggaaggaa ggaaggaagg aaggaaggaa	40	
25	<210> 26		
	<211> 44		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
30	<223> Oligonucleótido sintético		
	<400> 26		
	ggaaggaagg aaggaaggaa ggaaggaagg aaggaaggaa gga	44	
	<210> 27		
	<211> 28		
	<212> ADN		
35	<213> Secuencia artificial		
	<220>		
	<223> Oligonucleótido sintético		
	<400> 27		
40	ggaaggaagg aaggaaggaa gggaggaa	28	
	<210> 28		
	<211> 12		
	<212> ADN		
	<213> Secuencia artificial		

ES 2 718 343 T3

	<220> <223> Oligonucleótido sintético	
	<400> 28 gaaagaatg aa 12	
5	<210> 29 <211> 12 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 29 gaaaggaagg aa 12	
15	<210> 30 <211> 16 <212> ADN <213> Secuencia artificial	
20	<220> <223> Oligonucleótido sintético	
	<400> 30 gaaagaaagg aaggaa 16	
	<210> 31 <211> 102 <212> ADN <213> Secuencia artificial	
25	<220> <223> Oligonucleótido sintético	
	<400> 31 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaagg 60	
	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga 102	
30	<210> 32 <211> 85 <212> ADN <213> Secuencia artificial	
35	<220> <223> Oligonucleótido sintético	
	<400> 32 gagaggaagg aaaggaagga aggaaggaag gaaggaagga aggaaggaag gaaagaatga 60	
	aggaaggaag gaaggaaggg aggga 85	
40	<210> 33 <211> 85 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	

ES 2 718 343 T3

	<400> 33	
	gagaggaagg aaaggaagga aggaaggaag gaaggaagaa aggaaggaag gaaagaatga	60
	aggaaggaag gaaggaaggg aggga	85
5	<210> 34 <211> 98 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
10	<400> 34	
	gagaggaagg aaagagagga aggaaagga ggaaggaagg aaggaagga gaaaggaagg	60
	aaggaaagaa tgaaggaagg aaggaagga gggagggga	98
15	<210> 35 <211> 106 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 35	
	gagaggaagg aaagagagga aggaaagga ggaaggaagg aaggaagga ggaaggaaga	60
	aaggaagga gaaagaatg aaggaagga ggaaggaagg gagggga	106
20	<210> 36 <211> 102 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
25	<400> 36	
	gagaggaagg aaagagagga aggaaagga ggaaggaagg aaggaagga ggaaggaaga	60
	aaggaagga gaaagaatg aaggaagga ggaagggagg ga	102
30	<210> 37 <211> 102 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 37	
	gagaggaagg aaagagagga aggaaagga ggaaggaagg aaggaagga ggaaggaaga	60
	aaggaagga agaatgaagg aaggaagga ggaagggagg ga	102
35	<210> 38 <211> 90 <212> ADN <213> Secuencia artificial	

ES 2 718 343 T3

	<220> <223> Oligonucleótido sintético	
	<400> 38 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaag	60
	aatgaaggaa ggaaggaagg aagggagggga	90
5	<210> 39 <211> 89 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 39 gagaggaagg aaaggaagga aggaaggaag gaaggaagga agaaaggaag gaaggaaga	60
	atgaaggaag gaaggaagga agggagggga	89
15	<210> 40 <211> 90 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 40 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aagaaaggaa ggaaggaag	60
20	aatgaaggaa ggaaggaagg aagggagggga	90
	<210> 41 <211> 98 <212> ADN <213> Secuencia artificial	
25	<220> <223> Oligonucleótido sintético	
	<400> 41 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaagg	60
	aaggaaggaa agaatgaagg aaggaaggaa gggagggga	98
30	<210> 42 <211> 89 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
35	<400> 42 gagaggaagg aaaggaagga aggaaggaag gaaggaagga aggaagaaag gaaggaagga	60
	aagaatgaag gaaggaagga agggagggga	89

ES 2 718 343 T3

	<210> 43 <211> 106 <212> ADN <213> Secuencia artificial	
5	<220> <223> Oligonucleótido sintético	
	<400> 43 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60
	aagaaaggaa ggaaggaag aatgaaggaa ggaaggaagg gaggga	106
10	<210> 44 <211> 110 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
15	<400> 44 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60
	aagaaaggaa ggaaggaag aatgaaggaa ggaaggaagg aagggagggg	110
20	<210> 45 <211> 98 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 45 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaagaaa gaaaggaagg	60
	aaggaaagaa tgaaggaagg aaggaaggaa gggagggg	98
25	<210> 46 <211> 94 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
	<400> 46 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaagaaa ggaaggaagg	60
	aaagaatgaa ggaaggaagg aaggaaggga gggg	94
35	<210> 47 <211> 94 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	

ES 2 718 343 T3

	<400> 47	gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aagaaagaaa ggaaggaagg	60
		aaagaatgaa ggaaggaagg aaggaaggga gggaa	94
5	<210> 48 <211> 106 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
10	<400> 48 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaaga	60	
	aaggaaggaa ggaagaatg aaggaaggaa ggaaggaagg gagggaa		106
	<210> 49 <211> 98 <212> ADN <213> Secuencia artificial		
15	<220> <223> Oligonucleótido sintético		
	<400> 49 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60	
	aaggaagaaa tgaaggaagg aaggaaggaa gggagggaa		98
20	<210> 50 <211> 114 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
25	<400> 50 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60	
	aaggaagaaa ggaaggaagg aaagaatgaa ggaaggaagg aaggaaggga gggaa		114
30	<210> 51 <211> 94 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
35	<400> 51 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa gaaaggaagg	60	
	aaggaagaaa tgaaggaagg aaggaaggga gggaa		94
	<210> 52 <211> 102 <212> ADN <213> Secuencia artificial		

ES 2 718 343 T3

	<220> <223> Oligonucleótido sintético	
	<400> 52 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60
	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102
5	<210> 53 <211> 106 <212> ADN <213> Secuencia artificial	
10	<220> <223> Oligonucleótido sintético	
	<400> 53 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60
	aaggaaggaa ggaagaatg aaggaaggaa ggaaggaagg gaggga	106
15	<210> 54 <211> 102 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 54 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa gaaagaaagg	60
20	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102
	<210> 55 <211> 93 <212> ADN <213> Secuencia artificial	
25	<220> <223> Oligonucleótido sintético	
	<400> 55 gagaggaagg aaaggaagga aggaaggaag gaaggaagga aggaagaaag gaaggaagga	60
	aagaatgaag gaaggaagga aggaagggag gga	93
30	<210> 56 <211> 90 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
35	<400> 56 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaagaaa ggaaggaagg	60
	aaagaatgaa ggaaggaagg aagggaggga	90

ES 2 718 343 T3

	<210> 57 <211> 81 <212> ADN <213> Secuencia artificial	
5	<220> <223> Oligonucleótido sintético	
	<400> 57 gagaggaagg aaaggaagga aggaaggaag gaagaaagga aggaaggaaa gaatgaagga	60
	aggaaggaag gaagggaggg a	81
10	<210> 58 <211> 106 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
15	<400> 58 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaagg	60
	aaggaaggaa agaatgaagg aaggaaggaa ggaaggaagg gagggg	106
20	<210> 59 <211> 102 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	
	<400> 59 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggga ggaagaaagg	60
	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102
25	<210> 60 <211> 102 <212> ADN <213> Secuencia artificial	
30	<220> <223> Oligonucleótido sintético	
	<400> 60 gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaagaaa ggaaggaagg	60
	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102
35	<210> 61 <211> 102 <212> ADN <213> Secuencia artificial	
	<220> <223> Oligonucleótido sintético	

ES 2 718 343 T3

	<400> 61		
	gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaagg	60	
	aagggaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102	
5	<210> 62 <211> 94 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
10	<400> 62		
	gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaaggaagg	60	
	aaagaatgaa ggaaggaagg aaggaaggaa gga	94	
15	<210> 63 <211> 82 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
20	<400> 63		
	gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaatg	60	
	aaggaaggaa ggaagggagg ga	82	
25	<210> 64 <211> 102 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
30	<400> 64		
	gagaggaagg aaagagagga aggaaaggaa ggaaggaagg aaggaaggaa ggaagaaaga	60	
	aaggaaggaa ggaagaatg aaggaaggaa ggaagggagg ga	102	
35	<210> 65 <211> 102 <212> ADN <213> Secuencia artificial		
	<220> <223> Oligonucleótido sintético		
	<400> 65		
	gagaggaagg aaagagagga agaaaaggaa ggaaggaagg aaggaaggaa ggaagaaagg	60	
	aaggaaggaa agaatgaagg aaggaaggaa ggaagggagg ga	102	

ES 2 718 343 T3

	<220>		
	<223> Oligonucleótido sintético		
	<400> 66		
	gagaggaagg aaagagagga aggaaggaag gaaggaagga aggaaggaag gaaggaagga	60	
	agaaaggaag gaaggaaga atgaaggaag gaaggaagga agggagga	109	
5	<210> 67		
	<211> 98		
	<212> ADN		
	<213> Secuencia artificial		
10	<220>		
	<223> Oligonucleótido sintético		
	<400> 67		
	gagaggaagg aaagagagga aggaaagga ggaaggaagg aaggaagga gaaaggaagg	60	
	aaggaaaga tgaaggaaga aaggaagga gggagga	98	
15	<210> 68		
	<211> 89		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Oligonucleótido sintético		
	<400> 68		
	gagaggaagg aaaggaagga aggaaggaag gaaggaagga aggaaggaag gaaggaaga	60	
20	atgaaggaag gaaggaagga agggagga	89	

REIVINDICACIONES

1. Un método para determinar si una muestra obtenida de un individuo comprende ácido nucleico que comprende un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos, que comprende analizar una muestra que comprende ácido nucleico del individuo para detectar la presencia en el locus DYX2 de los siguientes haplotipos en el bloque de haplotipos de riesgo de DCDC2 que comprenden rs33914824, rs807694, rs707864, rs10456301, rs16889066 y rs9379651: (a) CGCGAG en un gen DCDC2; o (b) GACGAG en un gen DCDC2; o (c) tanto CGCGAG como GACGAG en un gen DCDC2; en donde si la muestra comprende al menos uno de (a) - (c), la muestra comprende un haplotipo asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje en seres humanos.
2. El método de la reivindicación 1 que comprende, además, analizar una muestra que comprende ácido nucleico del individuo para determinar la presencia en el locus DYX2 de al menos uno de: (d) rs4504469 en un gen KIAA0319; (e) rs2038137 en un gen KIAA0319; (f) rs2143340 en un gen KIAA0319; (g) cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319; y (h) cualquier combinación de CGCGAG en un gen DCDC2; GACGAG en un gen DCDC2; tanto CGCGAG como GACGAG en un gen DCDC2; rs4504469 en un gen KIAA0319; rs2038137 en un gen KIAA0319; rs2143340 en un gen KIAA0319; y cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340 en un gen KIAA0319.
3. El método de la reivindicación 1 o la reivindicación 2, en donde:
- (a) la muestra se analiza para detectar al menos un marcador de (a), (b) y (c), y al menos un marcador de (d), (e), (f) y (g); y/o
- (b) la DA es discapacidad lectora (DL) o trastorno del lenguaje (TL) o tanto DL como TL; y/o
- (c) la muestra es sangre, células o tejido; y/o
- (d) el método comprende además analizar la muestra para determinar la presencia del alelo 5 del gen DCDC2, el alelo 6 del gen DCDC2, o tanto el alelo 5 del gen DCDC2 como el alelo 6 del gen DCDC2, por ejemplo, en donde el alelo 5 del gen DCDC2 comprende la SEQ ID NO:35 y el alelo 6 de DCDC2 comprende la SEQ ID NO:36; y/o
- (e) el análisis implica métodos mediados por hibridación, secuenciación o combinaciones de estos.
4. El método de cualquiera de las reivindicaciones precedentes para determinar si una muestra obtenida de un individuo comprende al menos un marcador asociado a discapacidad lectora (DL) y trastorno del lenguaje (TL) concomitantes en seres humanos, que comprende analizar la muestra para determinar la presencia de al menos un marcador de (a), (b) y (c), y al menos un marcador en el gen KIAA0319 en el locus DXY2 para un haplotipo asociado a la susceptibilidad de desarrollar TL en seres humanos, opcionalmente en donde:
- (a) dicho al menos un marcador de (a), (b) y (c) es CGCGAG o GACGAG, y dicho al menos un marcador en el gen KIAA0319 es rs4504469; rs2038137; o rs2143340; y/o
- (b) el método comprende además analizar la muestra para determinar la presencia del alelo 5 del gen DCDC2 en el locus DYX2, el alelo 6 del gen DCDC2 en el locus DYX2, o tanto el alelos 5 del gen DCDC2 en el locus DYX2 como el alelo 6 del gen DCDC2 en el locus DYX2.
5. El método de cualquiera de las reivindicaciones 1 a 3 para determinar si una muestra obtenida de un individuo comprende un marcador asociado a trastorno del lenguaje (TL) en seres humanos, que opcionalmente comprende además analizar la muestra para determinar la presencia de al menos uno de: rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; rs12193738; rs2817198; rs10456309; rs985080; rs1554690; rs2533096; rs6951437; rs344470; rs344468; rs807694; rs482700; rs7695228; rs1940309; rs505277; rs476739; rs867036; rs867035; rs2071674; rs7694946; rs4823324; y un marcador para al menos uno de los siguientes genes: NEK2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; y PGD.
6. El método de cualquiera de las reivindicaciones 1 a 3 para determinar si los ácidos nucleicos (ADN, ARN) en un individuo comprenden marcadores de haplotipos que interactúan de una manera sinérgica, lo que da como resultado un trastorno del aprendizaje (TA) en seres humanos, que opcionalmente comprende además analizar la muestra para determinar la presencia de al menos uno de: un alelo de riesgo de DCDC2 y al menos un haplotipo de riesgo de KIAA0319, en donde dicho al menos un alelo de riesgo de DCDC2 es el alelo 5 del gen DCDC2 en el locus DYX2 (SEQ ID NO: 35) o el alelo 6 del gen DCDC2 en el locus DYX2 (SEQ ID NO: 36), y al menos un haplotipo de riesgo de KIAA0319 es un haplotipo de una variante de KIAA0319 que comprende un SNP que es rs4504469; rs2038137; o rs2143340, y en donde si la muestra comprende al menos un haplotipo de riesgo de DCDC2 o al menos un alelo de riesgo de DCDC2 y al menos un haplotipo de riesgo de KIAA0319, los ácidos nucleicos comprenden marcadores de haplotipos que interactúan de forma sinérgica, lo que da como resultado un TA en seres humanos.
7. El método de cualquiera de las reivindicaciones 1 a 3 para identificar o ayudar a identificar a un individuo que corre el riesgo de desarrollar al menos una dificultad del aprendizaje (DA), opcionalmente en donde:

(a) el método comprende además la etapa de analizar una muestra obtenida dl individuo para determinar la presencia de al menos un haplotipo que se encuentre en el gen KIAA0319 dentro del locus DYX2; y/o

(b) el método comprende además la etapa de analizar una muestra obtenida del individuo para determinar la presencia de rs4504469, rs2038137, rs2143340, o cualquier combinación de estos, en un gen KIAA0319 dentro del locus DYX2; o cualquier combinación de los haplotipos en (a) y (b).

8. El método de cualquiera de las reivindicaciones 1 a 3 para identificar a un individuo que tiene, o es susceptible de desarrollar, una dificultad del aprendizaje (DA), que comprende además determinar si el ácido nucleico en la muestra comprende un haplotipo del gen KIAA0319 asociado a la susceptibilidad de desarrollar trastorno del lenguaje (TL), en donde el haplotipo del gen DCDC2 y el haplotipo del gen KIAA0319 interactúan de forma sinérgica para disminuir el rendimiento en la eliminación de fonemas y para producir un trastorno del aprendizaje (TA) en seres humanos, en donde si la muestra comprende ambos haplotipos, se identifica que el individuo tiene o es susceptible de desarrollar un TA.

9. Una matriz o micromatriz que comprende un soporte que tiene una pluralidad de regiones discretas, teniendo cada región discreta al menos un fragmento de ácido nucleico manchado en ella, en donde cada región comprende secuencias que son complementarias a los ácidos nucleicos que abarcan al menos un marcador asociado a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) en seres humanos, en donde dicho al menos un marcador es al menos uno de los haplotipos en el bloque de haplotipos de riesgo de DCDC2 que comprende rs33914824, rs807694, rs707864, rs10456301, rs16889066 y rs9379651 definidos en la reivindicación 1, y opcionalmente al menos un marcador en una variante del gen KIAA0319.

10. La matriz o micromatriz de la reivindicación 9, en donde la matriz o micromatriz comprende secuencias complementarias de al menos dos marcadores diferentes asociados a la susceptibilidad de desarrollar una dificultad de aprendizaje (DA), que comprende, por ejemplo, en al menos una región, secuencias que son complementarias de un marcador en una variante del gen DCDC2 asociada a la susceptibilidad de desarrollar una dificultad del aprendizaje (DA) y, en al menos una región diferente, secuencias que son complementarias a un marcador en una variante del gen KIAA0319 asociada a la susceptibilidad de desarrollar una DA.

11. La matriz o micromatriz de la reivindicación 9 o la reivindicación 10:

(a) que comprende además uno o más marcadores en un gen DCDC2 seleccionado de CGCGAG, GACGAG, alelo 5 de READ1 (SEQ ID NO: 35), alelo 6 de READ1 (SEQ ID NO: 36), o cualquier combinación de dos, tres o cuatro de CGCGAG, GACGAG, alelo 5 de READ1 (SEQ ID NO: 35) y alelo 6 de READ1 (SEQ ID NO: 36); y en donde dicho uno o más marcadores en un gen KIAA0319 se seleccionan de rs4504469, rs2038137, rs2143340, o cualquier combinación de dos o tres de rs4504469, rs2038137 y rs2143340; y/o

(b) en donde la DA es un trastorno del lenguaje (TL) en seres humanos, y que comprende además uno o más marcadores en un gen DCDC2 seleccionado de GACGAG, alelo 6 de READ1 (SEQ ID NO: 36), rs793845, rs2799373, rs793862, rs793834, rs2792682, rs807704, rs707864, rs807694, o cualquier combinación de estos; y en donde dicho uno o más marcadores en un gen KIAA0319 se seleccionan de rs12193738, rs2817198, rs10456309, o cualquier combinación de estos, por ejemplo, que comprende además uno o más fragmentos de ácido nucleico en el soporte que comprenden secuencias que son complementarias a secuencias genómicas que flanquean a uno o más marcadores asociados a la susceptibilidad de desarrollar TL en seres humanos, en donde dicho uno o más marcadores se seleccionan de: rs985080, rs1554690, rs2533096, rs6951437, rs344470, rs344468, rs482700, rs7695228, rs1940309, rs505277, rs476739, rs867036, rs867035, rs2071674, rs7694946, rs4823324, y marcadores para los siguientes genes: NEK2; DLEC1; NARS; IL411; PKD2; ATF5; NUP62; SIGLEC11; ACAN; y PGD; y/o

(c) en donde la DA es discapacidad lectora (DL) en seres humanos, y además comprende uno o marcadores en un gen DCDC2 seleccionado de CGCGAG y alelo 5 de READ1 (SEQ ID NO: 35), o ambos; y en donde dicho uno o más marcadores en KIAA0319 es rs10456309, por ejemplo, que comprende además uno o más fragmentos de ácido nucleico en el soporte que comprenden secuencias que son complementarias a secuencias genómicas que flanquean a uno o más marcadores asociados a la susceptibilidad de desarrollar DL en seres humanos, en donde dicho uno o más marcadores se seleccionan de rs180950, rs2590673, rs892100, rs1792745, rs12546767, rs12634033, rs892270, rs10887149, rs10041417, rs6792971, rs4725745, rs12444778, rs1444186, rs2294691, rs10456309, rs1562422, y al menos un marcador para cada uno de los siguientes genes: MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622; OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF; y/o

(d) que comprende además uno o más fragmentos de ácido nucleico manchados en el soporte que comprenden secuencias que son complementarias a las secuencias genómicas que flanquean a uno o más marcadores asociados a la susceptibilidad de desarrollar una DA en seres humanos, en donde dicho uno o más marcadores se seleccionan de rs12636438; rs1679255; rs9521789; rs1983931; rs9814232; rs7995158; rs6573225; rs4082518; rs442555; rs259521; rs482700; rs7695228; rs1940309; rs505277; rs476739; rs867036; rs867035; rs2071674; rs7694946; rs4823324; rs180950; rs2590673; rs892100; rs1792745; rs12546767; rs12634033; rs892270; rs10887149; rs10041417; rs6792971; rs12636438; rs1679255; rs9521789; rs476739; rs505277; rs482700; rs7695228; rs867036; rs867035; rs1940309; rs16889556; rs1047782; rs1530680; rs12667130; rs6965855; rs985080;

ES 2 718 343 T3

rs4726782; rs1718101; rs10487689; rs1918296; rs737533; rs793845; rs2799373; rs793862; rs793834; rs2792682; rs807704; rs707864; rs12193738; rs2817198; rs10456309; rs985080; rs1554690; rs2533096; rs6951437; rs344470; rs344468; rs4725745; rs12444778; rs1444186; rs2294691; rs10456309; rs1562422; rs807694; rs3756814; rs3777663; rs9295626; rs7763790; rs6935076; rs9348646; rs2328791; rs2328791; rs2817201; rs9295626; rs4576240; 5 rs17307478, rs9356939, rs7763790, rs6456621; rs6456624, rs6935076, rs2038137, rs3756821, rs1883593, rs3212236; rs3777663, rs3756814, rs6931809, rs6916186, rs6933328, rs17491647; rs2328791; rs33914824a; rs807694a; rs707864a; rs10456301a; rs16889066a; rs9379651a; rs2817201; rs9295626; rs10456309; rs4576240; rs17307478; rs9356939; rs7763790; rs6456621; rs3756821; rs1883593; rs3212236; rs2294691; rs3777663; 10 rs3756814; rs6931809; rs6916186; rs6933328; rs17491647; rs9348646; rs1562422 y marcadores de los siguientes genes: R5H2; OR5H6; RRAGA; OR6B3; UMOD; A26C1A; FAM29A; CHRNA1; IFIT5; LOC643905; K2; DLEC1; NARS; IL4I1; PKD2; ATF5; NUP62; SIGLEC11; ACAN; PGD; MAP4; OR2L8; CRYBA4; OR2T8; KIAA1622; OR2AK2; DHX30; GEMIN6; C20orf10; y PPIF.

FIG. 1

A.

N.º de unidad	Unidad de repetición 1	Unidad de repetición 2	Unidad de repetición 3	Unidad de repetición 4	Región constante	Unidad de repetición 5	Región constante
Secuencia	GAGAGGAAGGAAA	GGAA	GAAA	GGAA	GGAAAGAATGAA	GGAA	GGGA
Número presente	1-2	4-10	0-2	0-2	Constante	3-4	Constante

B.

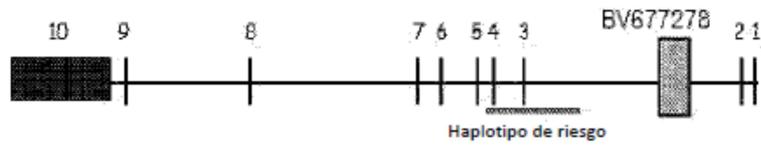
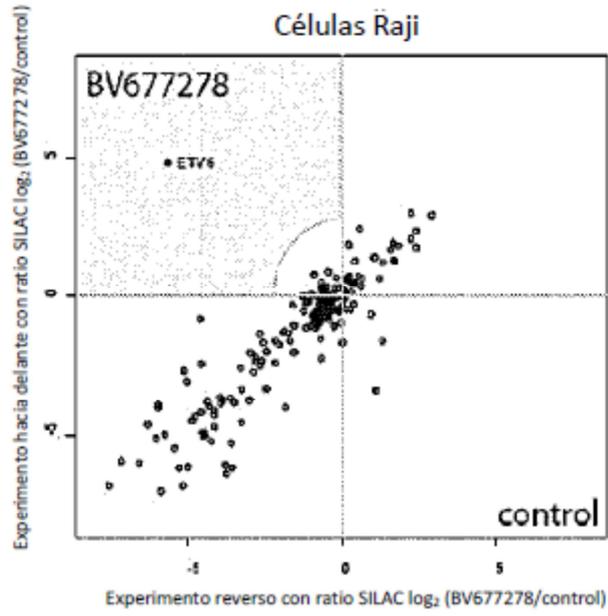


FIG. 2

A.



B.

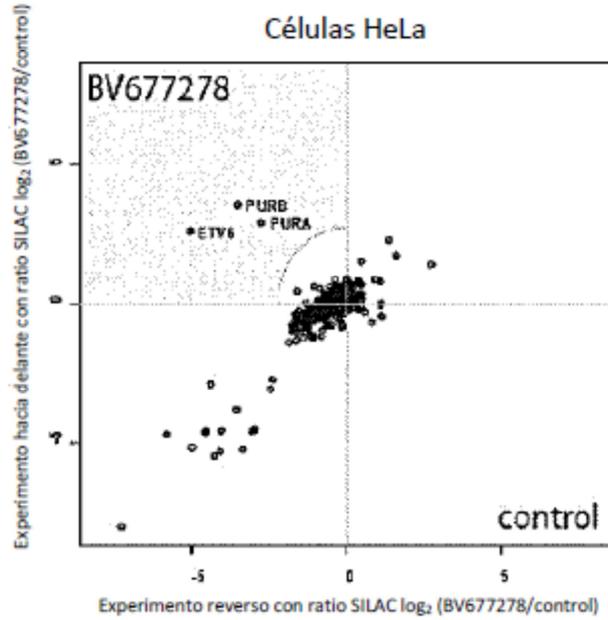


FIG. 2 Cont.

C.

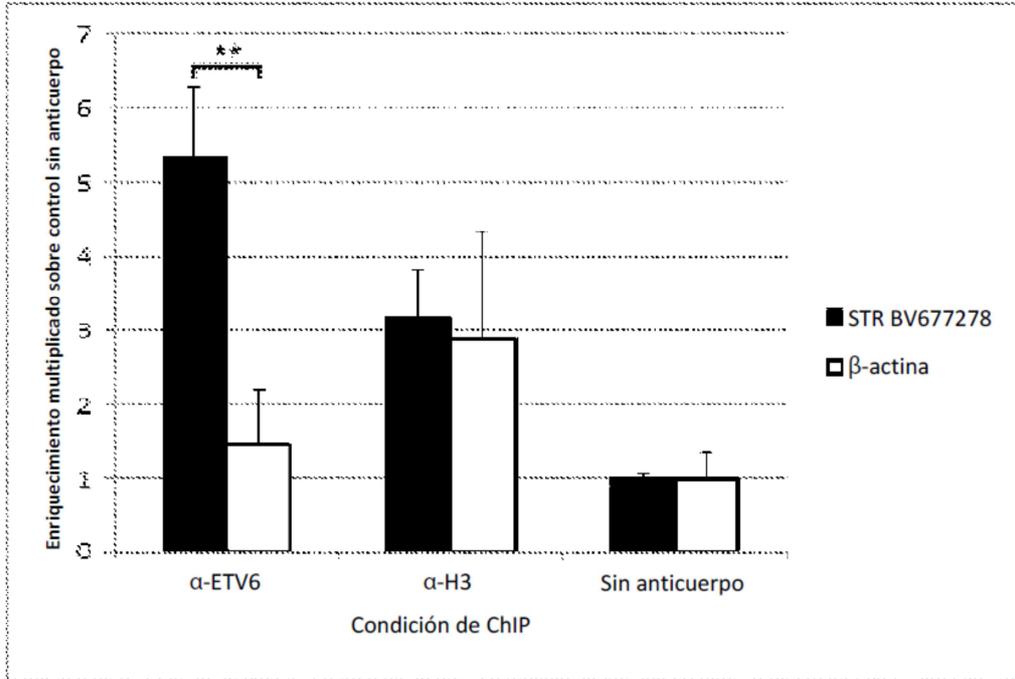
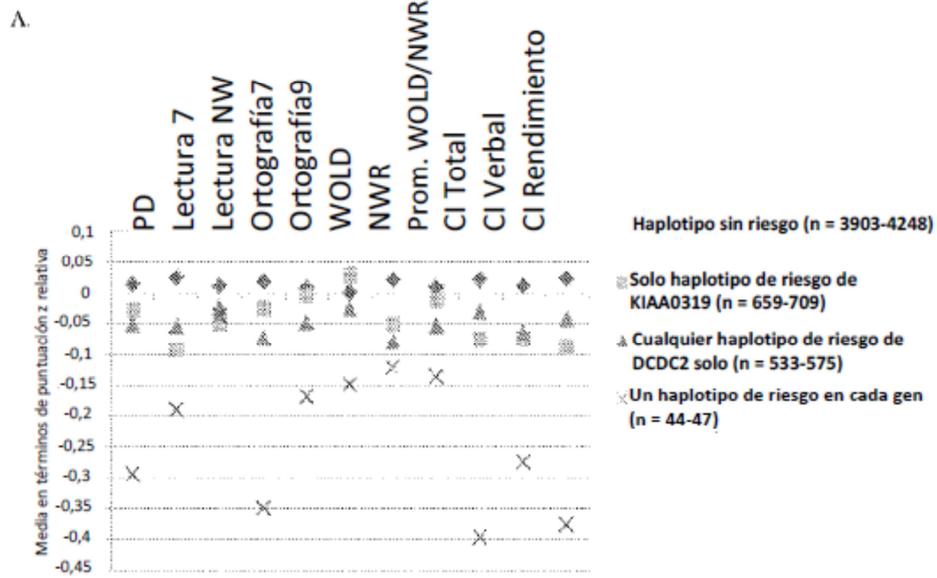


FIG. 3



B.

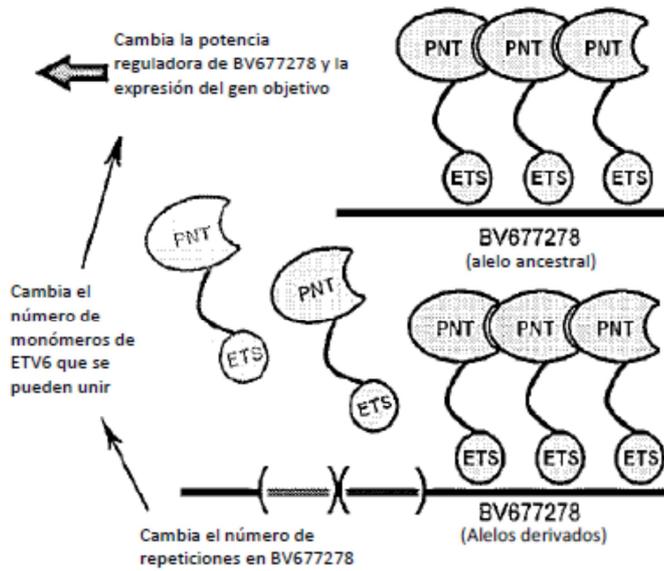


FIG. 4

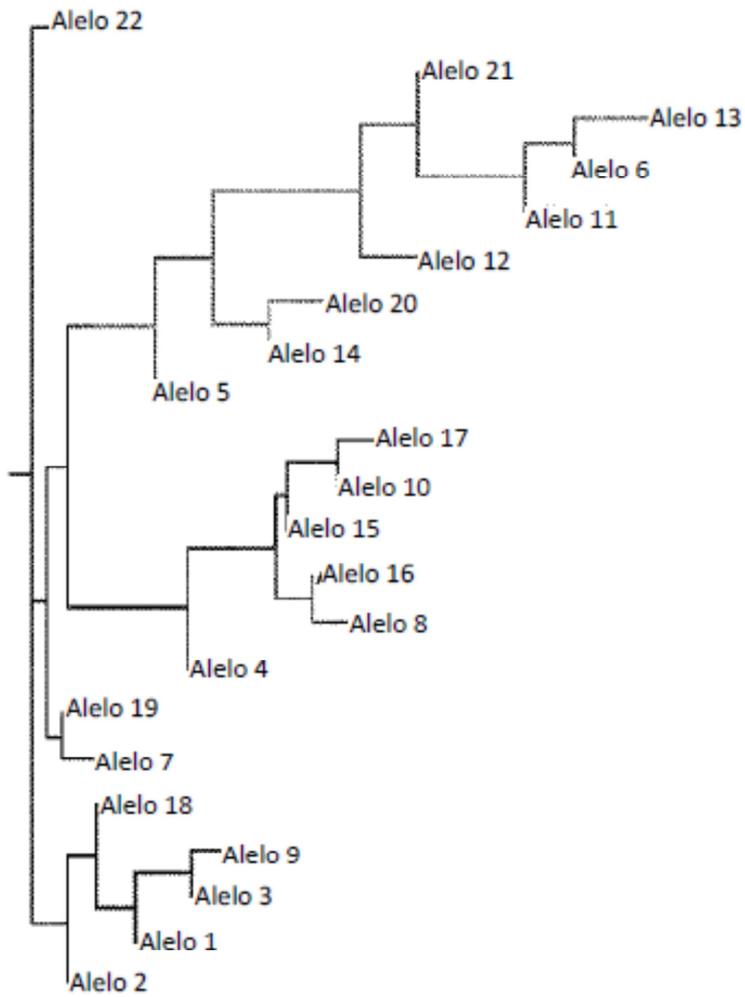


FIG. 5

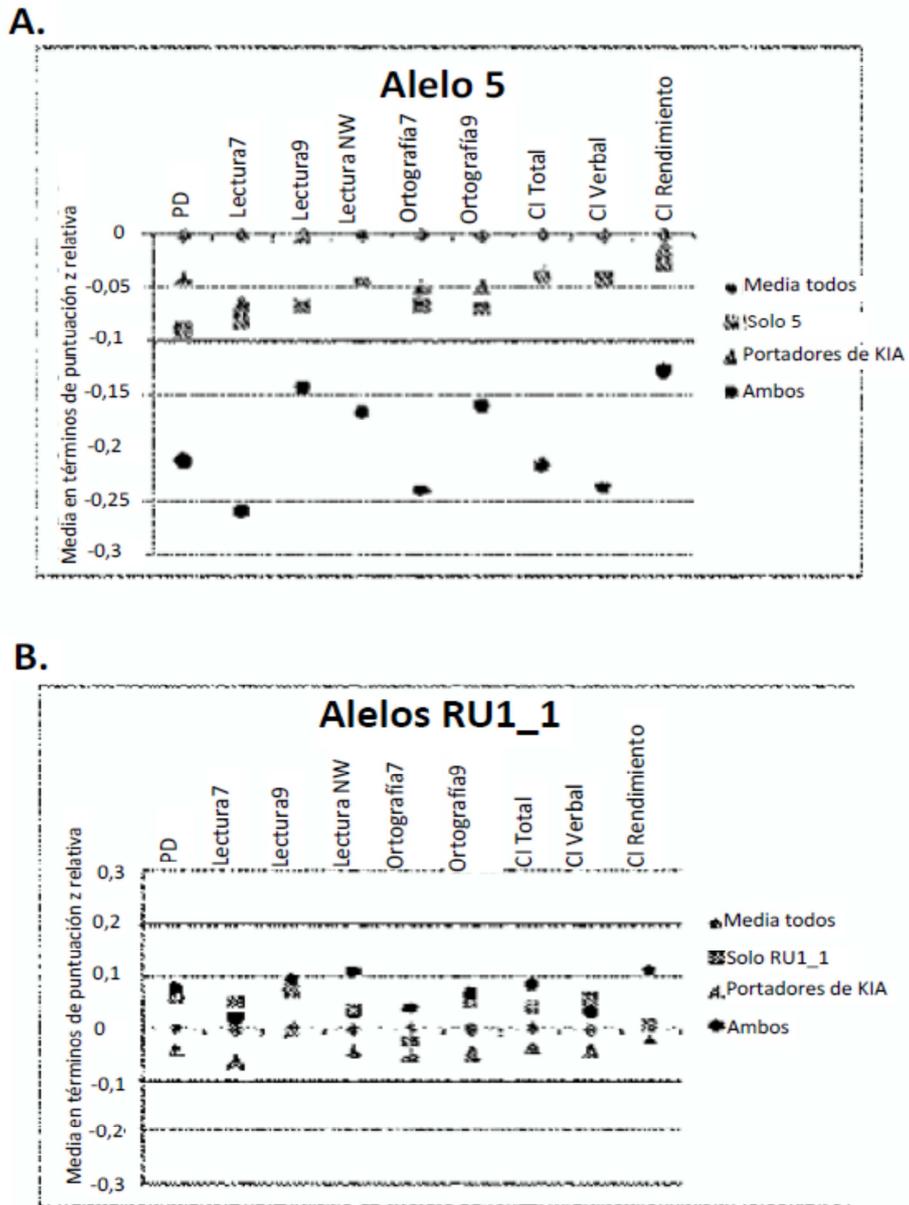
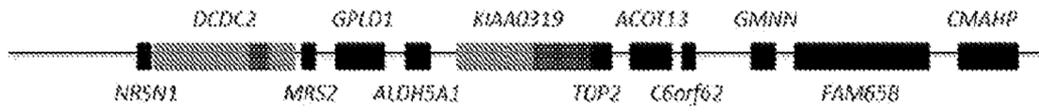


FIG. 6

A.



B.

